Published in Journals: Applied Sciences, Healthcare, Journal of Theoretical and Applied Electronic Commerce Research and Systems

Topic Reprint

Data Science and Intelligent Management

Edited by
Dongxiao Gu, Jiantong Zhang and Jia Li

mdpi.com/topics

MDPI

Data Science and Intelligent Management

Data Science and Intelligent Management

Topic Editors

Dongxiao Gu
Jiantong Zhang
Jia Li

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Topic Editors

Dongxiao Gu
School of Management
Hefei University of Technology
Hefei
China

Jiantong Zhang
School of Economics and Management
Tongji University
Shanghai
China

Jia Li
School of Business
East China University of Science and Technology
Shanghai
China

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Topic, published open access by the journals *Applied Sciences* (ISSN 2076-3417), *Healthcare* (ISSN 2227-9032), *Journal of Theoretical and Applied Electronic Commerce Research* (ISSN 0718-1876) and *Systems* (ISSN 2079-8954), freely accessible at: https://www.mdpi.com/topics/70E86291LA.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-8645-6 (Hbk)
ISBN 978-3-7258-8646-3 (PDF)
https://doi.org/10.3390/books978-3-7258-8646-3

© 2026 by the authors. Articles in this reprint are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The reprint as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

About the Editors . . . . . **ix**

Preface . . . . . **xi**

Xuan Liu, Xiaotong Chi, Jia Li, Shuqing Zhou and Yan Cheng
Doctors' Self-Presentation Strategies and the Effects on Patient Selection in Psychiatric Department from an Online Medical Platform: A Combined Perspective of Impression Management and Information Integration
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 13,
https://doi.org/10.3390/jtaer20010013 . . . . . **1**

Simona-Vasilica Oprea and Adela Bâra
Is Artificial Intelligence a Game-Changer in Steering E-Business into the Future? Uncovering Latent Topics with Probabilistic Generative Models
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 16,
https://doi.org/10.3390/jtaer20010016 . . . . . **27**

Xiaohong Miao, Zhongbin Li, Yingzheng Yan and Anxin Xu
A Tripartite Evolutionary Game-Based Cooperation Model of Cross-Border E-Commerce Logistics Alliances: A Case Study of China
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 37,
https://doi.org/10.3390/jtaer20010037 . . . . . **56**

Zhen Peng, Yunxiao Zhang and Tongtong Sun
Big Data-Driven Carbon Trading and Industrial Firm Value Based on DEA and DID
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 43,
https://doi.org/10.3390/jtaer20010043 . . . . . **77**

Mengyuan Peng, Kaixuan Zhu, Yadi Gu, Xuejie Yang, Kaixiang Su and Dongxiao Gu
How to Self-Disclose? The Impact of Patients' Linguistic Features on Doctors' Service Quality in Online Health Communities
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 56,
https://doi.org/10.3390/jtaer20020056 . . . . . **94**

Zicheng Zhang and Tianshu Zhang
Big-Data-Assisted Urban Governance: A Machine-Learning-Based Data Record Standard Scoring Method
Reprinted from: *Systems* **2025**, *13*, 320,
https://doi.org/10.3390/systems13050320 . . . . . **114**

Yuting Zhang and Jiantong Zhang
Spillover Effects of Physicians' Prosocial Behavior: The Role of Knowledge Sharing in Enhancing Paid Consultations Across Healthcare Networks
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 87,
https://doi.org/10.3390/jtaer20020087 . . . . . **141**

Kaihang Zhang, Changqi Dong, Yifeng Guo, Wuai Zhou, Guang Yu and Jianing Mi
Lagged Stance Interactions and Counter-Spiral of Silence: A Data-Driven Analysis and Agent-Based Modeling of Technical Public Opinion Events
Reprinted from: *Systems* **2025**, *13*, 417,
https://doi.org/10.3390/systems13060417 . . . . . **164**

Cong Liu, Yuan Shan and Jiming Cao
Effect of Network Structure on Conflict and Project Value Creation
Reprinted from: *Systems* **2025**, *13*, 594,
https://doi.org/10.3390/systems13070594 . . . . . **190**

Weitao Jiang, Hongxu Lu, Zexin Wang and Ying Jing
How Does Port Logistics Service Innovation Enhance Cross-Border e-Commerce Enterprise Performance? An Empirical Study in Ningbo-Zhoushan Port, China
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 188,
https://doi.org/10.3390/jtaer20030188 . . . . . **215**

Yunnan Cai, Jiangmin Chen and Shijie Li
Spatial Heterogeneity of Intra-Urban E-Commerce Demand and Its Retail-Delivery Interactions: Evidence from Waybill Big Data
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 190,
https://doi.org/10.3390/jtaer20030190 . . . . . **240**

Yu Wu, Qiao Lin, Jinming Wu, Ru Yao and Xuefu Zhang
Identification and Prediction Methods for Frontier Interdisciplinary Fields Integrating Large Language Models
Reprinted from: *Systems* **2025**, *13*, 677,
https://doi.org/10.3390/systems13080677 . . . . . **262**

Vlad Diaconita, Anda Belciu, Alexandra Maria Ioana Corbea and Iuliana Simonca
Decoding Wine Narratives with Hierarchical Attention: Classification, Visual Prompts, and Emerging E-Commerce Possibilities
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 212,
https://doi.org/10.3390/jtaer20030212 . . . . . **290**

Junchang Li, Jiaqing Sun and Jiantong Zhang
Dynamic Pricing for Internet Service Platforms with Initial Demand Constraints: Unified or Differentiated Pricing?
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 224,
https://doi.org/10.3390/jtaer20030224 . . . . . **329**

Wenlong Liu, Yashuo Yuan, Zifan Bai and Shenghui Sang
How Doctors' Proactive Crafting Behaviors Influence Performance Outcomes: Evidence from an Online Healthcare Platform
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 226,
https://doi.org/10.3390/jtaer20030226 . . . . . **366**

Yang Tong, Zexuan Shi and Jicai Li
Impacts of Brand Spillover Effect on Sourcing and Quality Disclosure of the Platform's Store Brand Under Asymmetric Information
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 291,
https://doi.org/10.3390/jtaer20040291 . . . . . **395**

Soheyla Tofighi, Faruk Gurbuz, Ricardo Mantilla and Shaoping Xiao
Advancing Machine Learning-Based Streamflow Prediction Through Event Greedy Selection, Asymmetric Loss Function, and Rainfall Forecasting Uncertainty
Reprinted from: *Appl. Sci.* **2025**, *15*, 11656,
https://doi.org/10.3390/app152111656 . . . . . **413**

Dongmei Han, Wangyan Jin, Zhengze Wu and Ruyi Ge
How Does Short Video Advertisement Congruence Drive Sales? The Underlying Mechanism of Sociability
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 312,
https://doi.org/10.3390/jtaer20040312 . . . . . **439**

Varadraj Gurupur, Sahar Hooshmand, Deepa Fernandes Prabhu, Elizabeth Trader and Sanket Salvi
Incompleteness of Electronic Health Records: An Impending Process Problem Within Healthcare
Reprinted from: *Healthcare* **2025**, *13*, 2900,
https://doi.org/10.3390/healthcare13222900 . . . . . **473**

Zhiyong Zhang, Liyan Xia, Yan Shi and Yongqiang Shi
A Four-Party Evolutionary Game Analysis of Silver Economy Data Sharing Based on Digital Platforms
Reprinted from: *Systems* **2026**, *14*, 27,
https://doi.org/10.3390/systems14010027 . . . . . **495**

Zhiwen Li, Jing Li, Miao Qi and Zhiwei Chen
Platform Data Transaction and Government Regulation Under Users' Personal Data Use Authorization Mechanism
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2026**, *21*, 12,
https://doi.org/10.3390/jtaer21010012 . . . . . **525**

Liu Fan, Shanshan Li, Can Wang and Xiaoping Zhang
Is a Chatbot More Effective? Investigating the Effect of Service Recovery Agents and Consumer Loss on Consumer Forgiveness
Reprinted from: *J. Theor. Appl. Electron. Commer. Res.* **2026**, *21*, 35,
https://doi.org/10.3390/jtaer21010035 . . . . . **547**

Lili Qu, Nan Hong and Yutong Han
Shipping Market Sentiment Shocks and BDI Volatility: Evidence from News-Based Indicators
Reprinted from: *Systems* **2026**, *14*, 317,
https://doi.org/10.3390/systems14030317 . . . . . **573**

Kaixuan Zhu, Lirong Song, Xuejie Yang, Wenxing Lu and Dongxiao Gu
Evolutionary Patterns and Advanced Strategies of Health Policies Based on Topic Modeling and Social Network Analysis
Reprinted from: *Systems* **2026**, *14*, 497,
https://doi.org/10.3390/systems14050497 . . . . . **592**

Zhengang Zhang and Defei Wang
The Impact of Digital Government on Regional Scientific and Technological Innovation Capacity
Reprinted from: *Systems* **2026**, *14*, 756,
https://doi.org/10.3390/systems14070756 . . . . . **617**

Leiming Gao, Ruixin Shi, Liuzi Wang, Shengzhi Jiao and Bei Wang
Identifying Key Predictors of Nursing Workload in Emergency Infusion Rooms: A Decision Tree Approach
Reprinted from: *Healthcare* **2026**, *14*, 1966,
https://doi.org/10.3390/healthcare14131966 . . . . . **652**

Arailym Medetbek, Aidos Mukhatayev, Svitlana Biloshchytska, Andrii Biloshchytskyi and Dias Berlibek
Machine Learning Analysis of Scientific Projects for Decision Support: Empirical Evidence from CORDIS FP6
Reprinted from: *Appl. Sci.* **2026**, *16*, 7056,
https://doi.org/10.3390/app16147056 . . . . . **667**

Ziyang Jin and Meiyan Li
Integrated Production and Transportation Scheduling in a Distributed Hybrid Flow Shop Considering Dynamic Orders Under an E-Commerce Environment
Reprinted from: *Appl. Sci.* **2026**, *16*, 7188,
https://doi.org/10.3390/app16147188 . . . . . **681**

About the Editors

Dongxiao Gu

Dongxiao Gu is a Professor at the School of Management, Hefei University of Technology, China. His research interests include business intelligence, smart healthcare, data analytics, and information systems. He has published extensively in leading international journals on topics related to online health communities, healthcare policy analysis, and intelligent decision-making. He serves as an associate editor and guest editor for several prominent journals in the fields of information systems and healthcare management. His current work focuses on applying machine learning and social network analysis to understand user behavior and optimize service quality in digital platforms.

Jiantong Zhang

Jiantong Zhang is a Professor at the School of Economics and Management, Tongji University, Shanghai, China. Her research focuses on digital economy, e-commerce, platform governance, and data-driven decision-making. She has authored numerous papers on online medical platforms, cross-border e-commerce, and the impact of digital technologies on business and society. She has served as a principal investigator for multiple research projects funded by the National Natural Science Foundation of China and other major funding agencies. Her recent work explores the intersection of artificial intelligence, consumer behavior, and platform dynamics.

Jia Li

Jia Li is a Professor at the School of Business, East China University of Science and Technology, Shanghai, China. His research spans behavioral science, electronic commerce, information systems, and data intelligence. He has published widely on topics including online consumer behavior, platform economics, and the application of big data analytics in management. He serves on the editorial boards of several international academic journals and has led numerous research projects in the areas of digital transformation and intelligent management. His current research examines how data-driven strategies can enhance organizational performance and user experience in digital ecosystems.

Preface

The rapid proliferation of big data and artificial intelligence has fundamentally reshaped the landscape of management research and practice. As organizations across sectors generate and collect unprecedented volumes of data, the ability to extract actionable intelligence from complex, high-dimensional datasets has become a critical competency. This Reprint, "Data Science and Intelligent Management," originates from a Topic that sought to capture the latest advances in applying data analytics to behavioral science, e-commerce, smart healthcare, and information systems. The 28 articles collected herein reflect the breadth and depth of contemporary research in this domain, spanning from theoretical modeling and algorithmic development to empirical investigations and policy analysis. The contributing authors employ a diverse array of methodologies—including machine learning, evolutionary game theory, social network analysis, topic modeling, and decision tree approaches—to address pressing challenges such as online patient–doctor interactions, cross-border e-commerce logistics, digital government transformation, platform data governance, and public health policy evolution. We are grateful to all the authors and reviewers whose rigorous work made this Reprint possible. It is our hope that this collection will serve as a valuable resource for researchers, practitioners, and policymakers seeking to harness the power of data science for intelligent management in an increasingly digital world.

Dongxiao Gu, Jiantong Zhang, and Jia Li
Topic Editors

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Doctors' Self-Presentation Strategies and the Effects on Patient Selection in Psychiatric Department from an Online Medical Platform: A Combined Perspective of Impression Management and Information Integration

Xuan Liu, Xiaotong Chi, Jia Li *, Shuqing Zhou and Yan Cheng

School of Business, East China University of Science and Technology, Shanghai 200237, China; xuanliu@ecust.edu.cn (X.L.); y10240182@mail.ecust.edu.cn (X.C.); zsqing1226@163.com (S.Z.); yancheng@ecust.edu.cn (Y.C.)
* Correspondence: jiali@ecust.edu.cn; Tel.: +86-21-64253177; Fax: +86-21-64253177

Abstract: Online medical consultation has become a crucial channel for patients seeking health support. Based on data from a psychiatric department in a leading online medical consultation platform in China, this study examines two possible types of online self-presentation strategies (positive impression management strategy and blending-in impression management strategy) employed by doctors in three dimensions: informational management, affective management, and image management, and explores their impact on patient selection. Meanwhile, an information integration perspective was incorporated and the interaction effects between impression management strategies taken by doctors and patient reviews expressed by online patients are also explored. Results indicate that the information quantity (representing the informational management dimension) in doctors' profiles has a negative impact on patient selection ($\beta = -0.142$, $p < 0.01$), while the positive emotion expression (representing affective management) ($\beta = 0.423$, $p < 0.01$) and profile photo (representing image management) ($\beta = 1.098$, $p < 0.01$) positively influence patient selection. Patient reviews related to expertise positively moderate the effect of information quantity in doctors' introduction on patient selection ($\beta = -0.632$, $p < 0.05$). In contrast, patient reviews concerning attitude ($\beta = -0.882$, $p < 0.01$) and credibility ($\beta = -0.488$, $p < 0.01$) negatively moderate the effect of emotion expression and profile photos on patient selection, respectively. The findings extend the applicability of impression management theory, providing a novel perspective for comprehending the impact of doctors' self-presentation on patient selection and its interaction effect with patient impressions.

Keywords: impression management; information disclosure; information integration; patient impressions

1. Introduction

1.1. Background and Research Questions

Online medical platforms have gradually become important channels for patients to enhance doctor–patient communication and seek medical support [1], while providing a new channel for doctor–patient interaction. In leading online medical platforms in China, such as Haodf Online (Haodf.com (accessed on 6 January 2025)) and Chunyu Doctor (Chunyuyisheng.com (accessed on 6 January 2025)), doctors provide online consultation services to patients so as to share medical and healthcare knowledge, and offer medical

support to community participants [2]. Patients seek health-related information support from online platforms to enhance their self-management and decision-making abilities [3]. This channel holds significant importance in alleviating the shortage of offline medical resources and improving the imbalance of medical resources. Doctors' participation in online medical services can significantly increase patient loyalty, improve doctors' performance, and enhance medical service efficiency [4].

Currently, online medical platforms are entering a stage of rapid development. From the perspective of the supply side of online consultation services, virtual communities have reduced the limitations of time and geography on medical services, leading to intensified competition among doctors [5]. According to data from the Haodf Online Platform, as of September 2024, the platform has over 280,000 real-name certified doctors providing services [6]. Among them, doctors with higher recognition and titles are more likely to attract more patients, leading to a clustering effect for top-tier doctors, also known as the Matthew effect on patient selection. In other words, approximately 80% of the orders are dominated by nearly 20% of doctors [7]. This poses challenges for ordinary doctors providing services on the platform. However, through effective impression management and information disclosure, ordinary doctors can also attract patient attention, influencing patient selections and helping to overcome the dilemma caused by the Matthew effect.

Impression management is the process by which individuals attempt to manage others' impressions of them. Depending on the audience, a person's style may differ [8]. Information disclosure and impression management are common concerns among online users [9]. When real-world interactions are taken to virtual spaces, social media users are encouraged to disclose a large amount of personal information in various forms, including text [10], labels [11], and photos [12–14]. Users on social media primarily manage their impressions and disclose information through positive strategies or blending-in strategies. Positive impression management strategies refer to users' efforts to portray themselves in an idealized manner by disclosing more information in order to maintain and build their reputations [15], while blending-in impression management strategies refer to users' efforts to hide unfavorable content and to reduce personal information disclosure [16].

However, online medical platforms differ from traditional social media. First, the trust mechanism on online medical platforms is distinct from the social trust mechanisms based on personal interests found on general social platforms. The former relies more on professional competence, certification information, and patient feedback. Second, the information on online medical platforms must not only be accurate but also comply with medical ethics and legal requirements. Consequently, the differences in impression management between doctors on online medical platforms and ordinary users on traditional social media are as follows: First, doctors manage their information disclosure on the platform carefully in an attempt to shape patients' impressions of their professionalism and abilities, ensuring they meet patient expectations to attract their selection. Second, in order to ensure the accuracy of disclosed information and compliance with rigorous platform regulations, doctors have less room for impression management. Therefore, this study thoroughly examines all dimensions of doctor impression management in detail and their impact on patient selection.

Doctor's impression management can be defined as the conscious display of their profile on online medical platforms, aiming to create a positive impression in patients' minds, build trust, and ultimately influence patient selection. In online medical platforms, doctor impression management can be managed in three ways: information management, affective management, and image management. (1) In terms of information management, doctors' personal information, such as education, specialization, experience in medicine, and other characteristics, is presented in their introductions to help patients understand

their professionalism [10]. To deepen patients' understanding of their doctors, doctors could use positive information strategies by providing richer descriptions of their past experience and professional expertise. However, due to the differences in professionalism between doctors and patients, excessive professional information may hinder the patient's understanding of the information [10]. Thus, to avoid misleading patients or increasing distance between doctors and patients, doctors could also implement the blending-in information strategy by simplifying and adjusting the content for easier comprehension. (2) As for affective management, doctors could convey emotion through their messages, expressing their attitude and offering encouragement and comfort to the patient. It is possible for doctors to use positive affective strategies to calm their patients' anxiety through positive emotions [17], reflecting their gentle personalities, which helps establish emotional connections and build trust, attracting patients' choice. However, emotional expressions may affect the doctor's professional image, especially when dealing with complex medical issues, where such expressions may seem lacking in rigor. Doctors can also adopt a blending-in emotion strategy to convey their professionalism and rigor through more conservative and serious words, but this strategy may not bridge the emotional gap with patients. (3) Regarding image management, doctors could choose a positive image strategy by displaying their real photo, which may enhance patient trust [18], or they can choose a blending-in image strategy by displaying cartoon avatars, thus reducing the influence of their appearance (such as attractiveness or dress) on patient decisions. Therefore, to explore the effectiveness of the two strategies in the three aspects of impression management, this study proposes the first research question:

Q1: *How do doctors' information, affective, and image strategies affect patient selection?*

Doctor–patient and patient–patient interaction is important on online medical platforms [17]. The process of doctor–patient interaction allows patients to have a deeper impression of the doctor. On online medical platforms, patients can publish their impressions of doctors based on their interactions with doctors. Previous studies focused on numerical types of patient feedback, such as voting and evaluation [19,20]. With the improvement of the feedback mechanism on online medical platforms, textual patient reviews that reflect patient impressions are also displayed on doctors' homepages. This information plays a crucial role in the patient selection process, as these patient impressions may enhance or weaken the effects of doctors' information disclosure and impression management. Specifically, these patient impression reviews can be categorized into expertise, attitude, and credibility types. Do these different types of patient impressions enhance or weaken the effectiveness of doctors' information management, affective management, and image management? When other patients see positive impressions of a doctor's expertise, attitude, and credibility, do they substitute for or enhance the effect of the doctor's information management, affective management, and image management? Does the aforementioned influence mechanism vary across different patient impression contexts? Therefore, this study further explores the second research question:

Q2: *What is the moderating role of different dimensions of patient impressions in doctors' impression management?*

The stigma faced by individuals with mental illness creates barriers to seeking care [21,22], making online consultations a more attractive option. In this context, impression management becomes crucial, as patients are particularly sensitive to the credibility and attitude of doctors, which can significantly influence their decision to seek treatment. Based on secondary data from the psychiatric department, this paper examines the impact

of two different strategies (positive impression management strategy vs. blending-in impression management strategy) on patient selection from three aspects of doctor impression management: information management (information quantity), affective management (emotion expression), and image management (profile photo). We also consider the moderating effects of patient impressions in textual reviews. Our results show that blending-in information management strategies, positive affective management strategies, and positive image management strategies help enhance patient selection. Patient impressions of doctors' expertise positively moderate the effect of information quantity on patient selection, while impressions regarding doctors' attitudes and credibility separately negatively moderate the effect of emotional expression and profile photos on patient selection.

Our research has several theoretical contributions. First, previous studies have explored patient selection from the perspectives of trust theory [23,24] and signaling theory [25]. These studies typically focus on basic information about doctors (such as titles, hospital levels, etc.) [24], or on the professional footprints of doctors (such as papers in scientific journals or presentations) [25]. In contrast, this study, based on impression management theory, comprehensively considers doctors' information management, affective management, and image management while controlling for basic information, providing a deeper analysis of doctor profiles and photos, and revealing the mechanisms that influence patient selection.

Second, while previous research mainly explored impression management on general social media, this study applies the positive and blending-in impression management strategies to online medical platforms. Our results show that in online medical platforms, these two strategies complement each other across different dimensions, revealing how doctors complement two strategies to increase patient selection, thus extending the understanding of impression management theory in the online medical field.

Third, we find that patient impressions of doctors' expertise amplify the negative effect of information quantity on patient selection, while impressions regarding doctors' attitude and credibility separately reduce the positive impact of emotional expression and profile photos on patient selection. These findings further explore the differential effects of doctors' impression management strategies in the presence of different types of patient impressions, extending the boundary conditions of impression management effectiveness.

This study provides practical guidance for doctors on information disclosure and impression management. It also offers valuable insights for online medical platforms on supporting ordinary doctors with these strategies and optimizing patient feedback display to help ordinary doctors attract patient selection.

1.2. Literature Review on Information Integration and Patient Selection Behavior

Information integration is the process by which an individual acquires, organizes, interprets, and integrates information from multiple sources to develop a comprehensive understanding of a particular topic and to make holistic judgments and decisions [26]. Information integration theory emphasizes how people bring together diverse information to form meaningful and coherent knowledge that leads to deeper understanding [27]. Information integration theory has been widely used in learning, judgment, attitude and impression formation, and decision making [28]. Wang et al. [29] argue that knowledge integration in online platforms has a direct positive effect on rational decision making, but an indirect negative effect on empirical decision making. Wong et al. [30] found that information integration between organizations affects customer service performance. Online medical platforms have provided new channels and conveniences for patients' medical counseling, and patients are becoming more active in their own health management. In this process, it is a combination of mass and interpersonal media. The Internet environment enhances patient interaction to meet their cognitive and emotional needs [31]. Multi-source

information is also integrated into patient care decisions in online medical platforms. Chen et al. [32] apply information integration to online platforms, suggesting that information from both doctors and patients is important during doctor–patient interactions. Patients need to improve their interaction skills and provide targeted resources to their doctors for more reliable online medical information.

Traditional social media focuses on broad, non-professional communication, whereas online medical platforms emphasize the doctor's professional identity, the platform's professionalism, and trust mechanisms. Therefore, many studies have examined doctors' performance from the perspective of doctor–patient interactions [2,17,33]. It has been found that characteristics such as emotion and professionalism displayed by doctors during consultation affect the information support and emotional support received by patients [32,34], which in turn affects patients' judgment and choice. With the development of various forms of information presentation by doctors in online medical platforms, multimodal information fusion, such as video, voice, and image, may have an impact on patient decision-making [35] and further stabilize the doctor–patient relationship. In terms of patient feedback information, factors such as patient ratings [10] and satisfaction [20,32] affect the establishment of patients' initial trust in their doctors. Doctors who receive more thank-you letters [20] and votes [32] have an elevated likelihood of their patients' selection, and platform resources are correspondingly skewed toward high-quality doctors [36,37].

Reviewing the existing literature reveals that most studies on doctors' personal performance focus on doctor–patient interactions. The impact of doctors' homepage information disclosure strategy on patient selection has been underappreciated. The doctor's homepage is the first channel for patients to learn about the doctor, which plays a crucial role in patient selection. Previous studies have identified the impact of patient feedback on medical selection. However, fewer studies have discussed the heterogeneous impact of different evaluation dimensions, such as efficacy and attitude, on patient selection. Finally, according to information integration theory, patients' choices are comprehensive judgments integrating multiple aspects of information. Therefore, it is vital to discuss the interaction between doctors' self-disclosure and patients' feedback.

1.3. Theoretical Foundation: Impression Management Theory

The concept of impression management was first introduced by Goffman [8], who argued that during an interaction between two people, it is common for an individual to inquire about the other's competence, attitude, and trustworthiness. The acquisition of this personal information can help determine each other's expectations for better interaction [8]. Impression management theory helps groups better understand citizens' roles and behaviors [38,39], and it is widely applied in enterprise management [40–43], interpersonal communication [44], and interview [45,46] contexts. With the continuous development of online platforms, more and more attention has been paid to the application of impression management in virtual communities [9]. Users on online platforms can easily manage their self-presentation, as the platform allows them to effortlessly create, modify, and edit information about themselves [16]. Delivering the right image to the right person can help users get the desired interpersonal results [47].

In virtual communities, users generally present themselves through online disclosure, and impression management based on disclosure becomes a tool for individuals or organizations to maintain their image [47–49]. Impression management strategies can be categorized based on different criteria. They may be role-related or task-related [50]. In the context of live-streaming e-commerce, these strategies can be further divided into self-focused, consumer-focused, and product-focused tactics. Self-focused and consumer-focused tactics help build trust in influencers, while consumer-focused and product-focused tactics foster

trust in products [51]. In interview scenarios, impression management strategies can be classified into honest and deceptive tactics [45]. In employee social contexts, they can be categorized into five types: ingratiation, exemplification, self-promotion, supplication, and intimidation, and it has been found that employees who avoid using the supplication tactic may face lower performance ratings [52]. Positive strategies and blending-in strategies represent two broad categories of impression management. In positive presentation strategies, users strive to present themselves in an idealized light and enhance their image [15]. Users are able to focus on and promote information that benefits them. They can also present themselves to others in a positive and socially desirable way to get closer to their idealized goals [53,54]. In blending-in strategies, users hide their personal characteristics and faulty aspects in order to blend in better with the group. When there is a discrepancy between an individual's desired self-image and their actual self-image or when the available information creates an impression barrier, individuals tend to delete unwanted information, such as managing the content of unwanted photographs, texts, and labels, and controlling the amount of self-disclosed information to avoid bad outcomes [11].

The disclosure of information on a doctor's homepage on an online medical platform helps patients generate a first impression of the doctor, which in turn influences patient selection [10,24]. However, most studies have considered only one or two aspects of information management, affective management, or impression management. The main components of impression management on doctors' homepages are information quantity, emotion expression, and profile photos. Doctors can choose positive or negative impression management strategies to present their information, affect, and image, respectively. Based on this, this paper explores the impact of impression management strategies on patients' selection from three aspects: information, affect, and image. On online medical platforms, in addition to doctors' own information disclosure, patients' feedback also impacts other patients' impressions of doctors [55]. Therefore, this paper considers the impact of the interaction between different impression management strategies and patients' feedback on patients' choices.

1.4. Research Hypotheses

1.4.1. Information Management

As part of the display of a doctor's personal information, the doctor can include a brief description of his education, practice experience, honors received, scientific research achievements, professional expertise, etc., to show how well he/she is qualified. Doctors' professionalism is associated with their patient's healthcare experience, so the textual expression of the doctor's professionalism influences patients' decisions regarding doctor selection. Doctors can choose positive information strategies to help patients reduce decision-making risk and uncertainty through a large amount of information, or they can choose blending-in information strategies to reduce information redundancy. It is the primary objective of patients to engage with online platforms in order to seek emotional and informational support [56]. Peng et al. [10] concluded that for patients, processing emotionally related information is a primary task, while processing other types of information, such as text, becomes secondary or lower-ranking. Processing textual information takes more time and effort from the patient. Excessive textual information may create barriers to patient comprehension and choice, so this paper argues that doctors should choose a blending-in information strategy to avoid information comprehension barriers, i.e., the information quantity of doctor's introduction has a negative effect on patient selection. Hence, we propose the following hypothesis:

H1: *In psychiatric departments, the doctors' information quantity in display negatively influences patient selection.*

1.4.2. Affective Management

Emotional expression plays an important role in managing impressions. Doctors can choose a positive affective strategy to show their optimistic and positive side through the expression of positive emotions; they can also choose a blending-in affective strategy that does not show too much emotion and conveys to patients a more serious message about their rigorous and serious attitude towards their work. Previous research has shown that positive emotion expression is positively correlated with user attention [57]. Individuals can make others feel positively affectionate towards them by expressing positive emotions [58]. As people want to be liked by others, they use a variety of strategies to express positive emotions [59]. The emotional expression of doctors on online medical platforms has been the focus of research. Past studies have shown that positive emotions displayed by doctors during doctor–patient communication can give more emotional support to patients [60], thus improving the treatment effect of diseases [61]. Doctors' presentation is a key channel for expressing positive emotions to potential patients. Positive emotions will provide emotional support and enhance patient intimacy, which in turn promotes patient selection. Therefore, this study argues that doctors should adopt positive affective strategies, and the positive emotional expression of doctors' messages can positively influence patients' selection. Hence, we put forward the following hypothesis:

H2: *In psychiatric departments, doctors' positive emotional expression on display can positively influence patient selection.*

1.4.3. Image Management

A doctor's avatar is the most visual representation of a their image and personal identity. Doctors can adopt positive image strategies to show their real avatars, or they can choose blending-in image strategies to publish cartoon avatars to hide their real characteristics. Past research has shown that authentic personal images are critical to the formation of personal impressions [62,63]. Users who post more photos on social media are more likely to be perceived as popular and approachable [64]. Crowdfunding performance is also positively affected by the number of images posted. Supporters can better understand and remember messages when images are used to convey them [12]. Consumers can be persuaded to trust and accept marketing products with more images [12]. A profile photo is also an important factor in online interactions. Real personal photography helps to carefully construct an online identity [18]. On online medical platforms, patients can get an impression of their doctors through doctor images. Patients will feel more comfortable if they see a real doctor's image. Moreover, a real image will reduce the distance between patient and doctor and enhance their trust. Therefore, this study argues that doctors should choose a positive image strategy of posting real avatars, i.e., doctors uploading real images will increase their patients' selection. Hence, we put forward the following hypothesis:

H3: *In psychiatric departments, doctors' use of profile photos on display can positively influence patient selection.*

1.4.4. Moderating Effects of Patient Impressions

According to information integration theory, patients are not only recipients of information from doctors, but also from other patients. Patients can access information disclosed by doctors themselves, as well as referring to unfiltered impressions of other

patients. Patient impressions can convey additional information about the doctor and the services to potential patients, reducing uncertainty and the risk of patient decision [65]. Indeed, online patient impressions are often considered to be more reliable and influential [66,67]. Cheng and Ho [68] find that both positive and negative consumer feedback are perceived by potential consumers as effective signals to reduce uncertainty and can greatly influence their purchase decisions. Liu [67] finds that product sales are positively affected by the amount, and potency (positive or negative) of online feedback from users. On online medical platforms, positive feedback benefits doctors' reputations, and patients prefer to share treatment experiences for doctors who have a higher online reputation [69]. As a kind of patient information disclosure, patient impressions can supplement the information disclosed by doctors and are not controlled by doctors' subjective information presentation strategies. Thus, this paper further considers the moderating role of patient impressions in the relationship between doctors' disclosures and patient selection. Patient impressions were categorized into three dimensions: expertise, attitude, and credibility, corresponding to information, emotion, and image.

In terms of information, we argue that patient impressions of doctors' expertise preference are a substitute for the information presented in doctor introductions. When the patient's impression of the doctor's expertise preference is high, it increases information support and the patient's reliance on the doctor's introduction decreases. At this point, if the doctor displays a lengthy textual introduction, it will cause patient boredom, and the negative impact on patient selection will be greater. Hence, we put forward the following hypothesis:

H4a: *In psychiatric departments, patient impressions of doctors' expertise preference positively moderate the negative effect of information quantity on patient selection.*

In terms of the affective dimension, we believe that a patient's impression of doctors' attitude preferences plays an alternative role to a doctor's emotional expression in their message. Emotional support will increase when a patient's impression of the doctor's attitude preference is high. Patient impressions also make it easier for other patients to visualize the positive attitude of the doctor. At this point the positive role of emotional expression in the doctor's message is reduced. Hence, we put forward the following hypothesis:

H4b: *In psychiatric departments, patient impressions of doctors' attitudes preference negatively modulate the positive effects of emotional expression on patient selection.*

In terms of image dimension, we similarly consider patient impressions of doctors' credibility preference to substitute doctor disclosure. A patient's impression of a doctor's credibility preference is based on other patients' perspectives, which can enhance the doctor's sense of authenticity and reliability. At this point, the positive effect of the doctor's real image on patient selection diminishes. Hence, we put forward the following hypothesis:

H4c: *In psychiatric departments, patient impressions of doctors' credibility preference negatively moderate the positive effect of profile photo use on patient selection.*

The conceptual model for this study is shown in Figure 1.

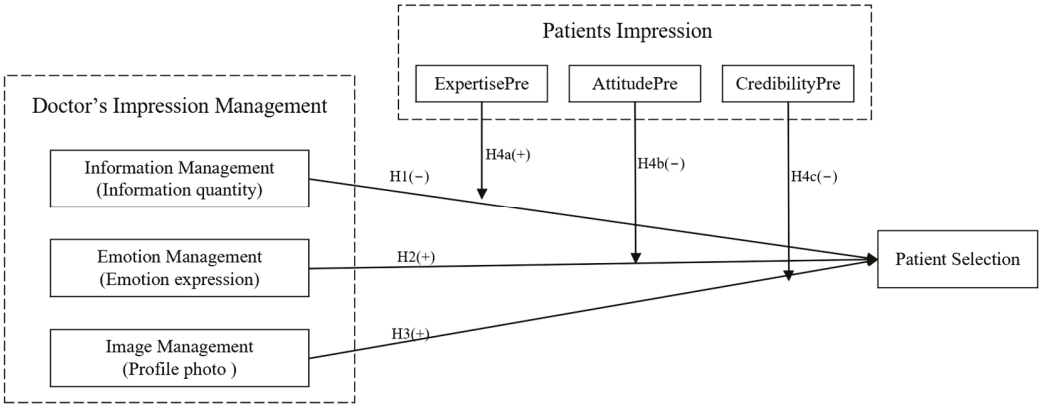


Figure 1. Conceptual model.

2. Materials and Methods

2.1. Data Collection

In this paper, data is taken from the leading Internet platform in China, Haodf.com. As of September 2024, Haodf.com has over 280,000 real-name certified doctors providing services and has served more than 89 million patients [6], making it a comprehensive and representative source for this study. Stigma and discrimination against individuals with mental illness remain a persistent concern when seeking medical help [21,22], making them more likely to seek assistance in an online setting rather than opting for face-to-face consultations. Moreover, due to the nature of mental illnesses, patients are often more sensitive to the credibility and attitude of doctors. Therefore, the psychiatric department is particularly suitable for our empirical analysis in the context of online doctor–patient interactions. We firstly obtain the data of all psychiatrists based on the list of doctors and further obtain the data from each doctor's homepage. The information includes the number of patient selections, doctor's photo, doctor's profile, doctor's message, patient impression, doctor's hospital level, hospital label, doctor's title, hot, honors, reply rate, services, and homepage visits. After excluding records with missing values, we get a total of 1036 doctors for analysis. An example of a doctor's homepage is shown in Figure 2.

Figure 2. Doctor's homepage.

2.2. Variables

(1) Dependent variable

The dependent variable of this study is *Patient selection*, as this decision typically occurs in the initial stage of a patient's consultation and is highly influenced by doctors' impression management. Patient selection can be measured by the number of patients served by the doctor. On an online medical platform, this is calculated by a doctor's online consultation volume (log-transformed).

(2) Independent variables

The first independent variable is *Information quantity*, *which refers to the amount of information* [70], typically measured by the length of the information [33,71]. Initially, we obtain textual information from a doctor's homepage introduction. Subsequently, we employ the "Jieba" Chinese word segmentation module [34] in Python to segment the text, eliminate non-contributory words, and quantify the remaining words to assess the information quantity of a doctor on display.

The second independent variable is *Emotion expression*. Initially, we retrieved the textual content from the doctor's messages. Subsequently, we assessed the doctor's emotional disposition using SnowNLP, a Python library for Chinese text processing and sentiment analysis [17,72], to score the emotional inclination within the doctor's messages.

The third independent variable is *Profile photo*. Previous studies have used the criterion "whether the doctor provides photographs on the platform" [24]. In our study, we code this based on whether the doctor's avatar on the homepage is a real photo or not, 1 refers to real photo, 0 otherwise.

(3) Moderation variables

After selecting a doctor's service, patients can evaluate him/her to show the impression the doctor left. We categorized all doctors' impressions based on the polarity (positive vs. negative) and the relationship between the evaluation content and the three dimensions of the independent variables. Specifically, expertise, attitude, and credibility correspond to information management, affective management, and image management, respectively.

We initially randomly selected 100 patient reviews, which were manually classified by two investigators. They determined the type of each review as either expertise, attitude, or credibility and annotated the impression polarity. Positive impressions were marked as 1, while negative impressions were marked as -1. Based on these results, we developed classification guidelines. In these guidelines, expertise refers to the ability to solve problems, attitude refers to the service attitude during the service process, and credibility refers to being trustworthy or responsible. Examples are provided in Table 1. The two investigators shared these guidelines, and we measured intercoder reliability using the agreement rate, which was 0.930, exceeding 0.9. To resolve any inconsistencies in the coding process, a third investigator was invited to review the discrepancies.

Table 1. Patient impressions are categorized according to different elements.

Polarity	Expertise	Attitude	Credibility
Positive impression	Conceptual clarity Exceptional medical expertise Outstanding medical skills	Good service attitude Service with a smile Genial	Trustworthy Help patients find hope Convincing
Negative impression	No explanation of the problem Lack of detail in consultation If there is a problem, doctors are unconcerned	Reprimanding the patient with anger Voice of impatience Rush the patient out.	Not yet available

Based on the patient impressions displayed on each doctor's homepage, we calculated three moderating variables, *ExpertisePre*, *AttitudePre*, and *CredibilityPre* using Formulas (1)–(3).

$$\textbf{\textit{ExpertisePre}} = (Number\ of\ expertise\ related\ (positive)\ tags - Number\ of\ expertise\ related\ (negative)\ tags)\ /\ Total\ number\ of\ expertise\ related\ tags \quad (1)$$

$$\textbf{\textit{AttitudePre}} = (Number\ of\ attitude\ related\ (positive)\ tags - Number\ of\ attitude\ related\ (negative)\ tags)\ /\ Total\ number\ of\ attitude\ related\ tags \quad (2)$$

$$\textbf{\textit{CredibilityPre}} = (Number\ of\ credibility\ related\ (positive)\ tags - Number\ of\ credibility\ related\ (negative)\ tags)/\ Total\ number\ of\ credibility\ related\ tags \quad (3)$$

(4) Control variables

In this study, HospitalRank, HospitalTitle, DoctorTitle, Hot, Honor, Reply, Services and Visits were selected as control variables. HospitalRank [7,24] and HospitalTitle describe the rank and label of a doctor's hospital. These two hospital-level variables may influence patients' perceptions of doctor competence. DoctorTitle reflects the doctor's education and qualifications [2,24]. Hot is given by the platform's calculations, which affect the doctor's position on the doctor's list [32] and influence patients' selection. Honor refers to the number of consecutive "Good Doctor of the Year" awards a doctor has received, controlling for this variable because patients generally trust doctors with more honors. Reply refers to the response rate at which a doctor responds to a patient. This influences the frequency of doctor–patient interactions, with more frequent interactions potentially having a positive impact on patient selection. Services refers to the number of services a doctor provides, which expresses the diversity of a doctor's services and may affect the online consultations. Visits is the number of people who visit the doctor's homepage. It is included to control for the effect of patient traffic, since, ultimately, patient selection is converted by visits to the doctor's homepage.

Table 2 summarizes the definitions of all variables.

Table 2. Variable Definition.

Variable Type	Variable Name	Variable Definition
Dependent Variable	Patient selection	The number of online doctor consultations (log-transformed)
Independent Variables	Information quantity	The number of words included in the doctor's presentation (log-transformed)
	Emotion expression	Emotional expression in doctor's messages
	Profile photo	Whether the doctor's avatar on the homepage is a real photo or not, the real photo is coded as 1, and 0 otherwise
Moderator variables	ExpertisePre	Patient impressions (doctor's professional relevance)
	AttitudePre	Patient impression (doctor's attitude related)
	CredibilityPre	Patient impressions (doctor trust related)
Control variables	HospitalRank	Rank of the doctor's hospital, with 1–4 being ranked II, IIA, III, and IIIA, respectively
	HospitalTitle	Whether the doctor's hospital is a psychiatric hospital, with 1 representing a psychiatric hospital and 0 representing otherwise
	DoctorTitle	Doctors' title, reflecting the doctor's education and qualifications, 1–4 are junior, intermediate, associate senior, and full senior doctors, respectively.
	Hot	Hot value of doctors on the platform (Generated by platform)
	Honor	Number of times a doctor was awarded "Good Doctor of the Year"
	Reply	The speed of the doctor's response. Labeled 1–5 from slowest to fastest
	Visits	Number of visitors to the doctor's homepage (log-transformed)
	Services	Number of different services opened by doctors

2.3. Methods

Firstly, this study conducts descriptive statistics on all variables, then calculates Spearman correlation coefficients to address potential multicollinearity issues. Finally, to test the hypotheses, this study performs Ordinary Least Squares (OLS) regression analysis. Four models are constructed in this paper, including one main effect model and three moderating effect models. Among them, the main effect model is shown in Equation (5). The moderating effect models of patient impression on *Information quantity*, *Emotion expression* and *Profile photo* are shown in Equations (6), (7) and (8), respectively.

$$Y_i = \sigma_0 + \sigma_1 HospitalRank_i + \sigma_2 HospitalTitle_i + \sigma_3 DoctorTitle_i + \sigma_4 Hot_i + \sigma_5 Honor_i + \sigma_6 \text{Reply}_i + \sigma_7 \text{Invisits}_i + \sigma_8 Services_i + \varphi_i \tag{4}$$

$$Y_i = \alpha_0 + \alpha_1 HospitalRank_i + \alpha_2 HospitalTitle_i + \alpha_3 DoctorTitle_i + \alpha_4 Hot_i + \alpha_5 Honor_i + \alpha_6 \text{Reply}_i + \alpha_7 \text{Invisits}_i + \alpha_8 Services_i + \alpha_9 Information\ quantity_i + \alpha_{10} Emotion\ expression_i + \alpha_{11} Profile\ photo_i + \vartheta_i \tag{5}$$

$$Y_i = \beta_0 + \beta_1 HospitalRank_i + \beta_2 HospitalTitle_i + \beta_3 DoctorTitle_i + \beta_4 Hot_i + \beta_5 Honor_i + \beta_6 \text{Reply}_i + \beta_7 \text{Invisits}_i + \beta_8 Services_i + \beta_9 Information\ quantity_i + \beta_{10} Emotion\ expression_i + \beta_{11} Profile\ photo_i + \beta_{12} ExpertisePre_i + \beta_{13} Information\ quantity_i * ExpertisePre_i + \varepsilon_i \tag{6}$$

$$Y_i = \gamma_0 + \gamma_1 HospitalRank_i + \gamma_2 HospitalTitle_i + \gamma_3 DoctorTitle_i + \gamma_4 Hot_i + \alpha\gamma_5 Honor_i + \gamma_6 \text{Reply}_i + \gamma_7 \text{Invisits}_i + \gamma_8 Services_i + \gamma_9 Information\ quantity_i + \gamma_{10} Emotion\ expression_i + \gamma_{11} Profile\ photo_i + \gamma_{12} AttitudePre_i + \gamma_{13} Emotion\ experssion_i * AttitudePre_i + \epsilon_i \tag{7}$$

$$Y_i = \delta_0 + \delta_1 HospitalRank_i + \delta_2 HospitalTitle_i + \delta_3 DoctorTitle_i + \delta_4 Hot_i + \delta_5 Honor_i + \delta_6 \text{Reply}_i + \delta_7 \text{Invisits}_i + \delta_8 Services_i + \delta_9 Information\ quantity_i + \delta_{10} Emotion\ expression\ _i + \delta_{11} Profile\ photo_i + \delta_{12} CredibilityPre_i + \delta_{13} Profile\ photo_i * CredibilityPre_i + \theta_i \tag{8}$$

2.4. Ethical Cautions

Ethical review and approval were waived for this study, due to the data used being publicly available from Haodf.com and not involving any patient privacy information.

3. Results

3.1. Descriptive Statistics

Table 3 indicates descriptive statistics for all variables. From the results in the table, it can be seen that there is a large range when it comes to the popularity of doctors (*Patient selection*) and the number of homepage visits (*Visits*). As for independent variables, the mean for *Information quantity* is 3.9, indicating that the average number of words in a doctor's profile is 48.4. The mean for *Emotion expression* is 0.35, suggesting that doctors' messages are predominantly negative in tone. The mean for *Profile photo* is 0.57, meaning that more than half of the doctors in the study use a real photo as the profile picture on their homepage. In terms of patient impressions, the means for *ExpertisePre*, *AttitudePre*, and *CredibilityPre* are 0.09, 0.22, and 0.35, respectively. The average hospital level of the doctors is relatively high, with a mean of 3.83, and 68% of the doctors are from psychiatric specialty hospitals. Doctors with high titles are in the majority, but those awarded "Good Doctor of the Year" are relatively few. The doctors have a high hot score, with an average of 3.39. The average response rate of the doctors is greater than 4 (faster), and most of the doctors are responding at a faster rate or more, keeping it at a more active level. The average number of services provided by doctors is less than two, indicating most doctors still have a single type of service, and most have only offered the most basic services.

Table 3. Descriptive statistics.

Variable	Observations	Mean	Std. Dev.	Min	Max
Patient selection	1036	4.198	2.603	0	10.043
Information quantity	1036	3.90	1.17	0.00	7.08
Emotion expression	1036	0.35	0.46	0.00	1.00
Profile photo	1036	0.57	0.50	0.00	1.00
ExpertisePre	1036	0.09	0.14	−0.40	1.00
AttitudePre	1036	0.22	0.29	−1.00	1.00
CredibilityPre	1036	0.35	0.48	−1.00	1.00
HospitalRank	1036	3.83	0.67	0.00	4.00
HospitalTitle	1036	0.68	0.47	0.00	1.00
DoctorTitle	1036	3.10	0.91	0.00	4.00
Hot	1036	3.39	0.28	3.00	5.00
Honor	1036	0.05	0.39	0.00	7.00
Reply	1036	4.02	0.84	1.00	5.00
Services	1036	1.48	1.07	0.00	5.00
Visits	1036	11.48	2.13	4.08	17.84

3.2. Correlation Analysis

Table 4 demonstrates the correlations between the variables. According to the results in the table, *Invisits* shows a high correlation with *Patient selection*, with a coefficient of 0.813. This is because a higher number of patients visiting a doctor's homepage provides more opportunities for the doctor to be selected. There are correlation coefficients of 0.949, 0.886, and 0.855 between the three types of patient impressions. However, the empirical evidence does not create a multicollinearity problem because the three variables moderate different variables and do not co-exist in the same model. Furthermore, the correlation coefficients between the rest of the variables are less than 0.6, indicating that there is no serious problem with multicollinearity.

3.3. Regression Results

Table 5 shows the regression results with only control variables added to model (1). The results show that *Hot* (β = 0.646, p < 0.01), *Service* (β = 0.539, p < 0.01) and *Visits* (β = 0.885, p < 0.01) all have positive effects on patient selection, while *HospitalRank* (β = −0.31, p < 0.01), *DoctorTitle* (β = −0.253, p < 0.01) and *Honor* (β = −0.293, p < 0.05) all have negative effects on patient selection. With high hot ratings, doctors are more likely to be seen, and patients are more likely to select them. Doctors who offer a wide range of services serve a wide population, so their diversity positively influences patient selection. More visits to the doctor's homepage indicate that more attention is received, and accordingly, patient selection is increasing. The regression results about the hospital labels show that patients prefer to choose general hospitals rather than to be treated in specialized hospitals. It is likely that doctors with high titles and many honors, mostly well-known specialists with a lot of experience, will not put all their effort into operating the online platform, which will result in a lower online patient selection.

Table 4. Correlation analysis.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) Patient selection	1														
(2) Information quantity	0.356 ***	1													
(3) Emotion expression	0.477 ***	0.276 ***	1												
(4) Profile photo	0.518 ***	0.259 ***	0.346 ***	1											
(5) ExpertisePre	0.467 ***	0.125 ***	0.216 ***	0.336 ***	1										
(6) AttitudePre	0.520 ***	0.146 ***	0.238 ***	0.365 ***	0.855 ***	1									
(7) CredibilityPre	0.531 ***	0.138 ***	0.242 ***	0.369 ***	0.886 ***	0.949 ***	1								
(8) HospitalRank	−0.102 ***	−0.00300	−0.074 **	−0.056 *	−0.087 ***	−0.076 **	−0.098 ***	1							
(9) HospitalTitle	−0.147 ***	−0.200 ***	−0.00500	−0.127 ***	−0.132 ***	−0.140 ***	−0.135 ***	0.086 ***	1						
(10) DoctorTitle	0.268 ***	0.526 ***	0.148 ***	0.055 *	0.085 ***	0.096 ***	0.090 ***	−0.063 **	−0.0210	1					
(11) Hot	0.414 ***	0.213 ***	0.203 ***	0.178 ***	0.273 ***	0.311 ***	0.295 ***	0.0220	−0.0240	0.272 ***	1				
(12) Honor	0.200 ***	0.0470	0.119 ***	0.101 ***	0.108 ***	0.107 ***	0.093 ***	−0.088 ***	0.0160	0.0470	0.351 ***	1			
(13) Reply	−0.025	−0.080 **	0.0400	0.0130	0	0.00500	−0.00300	−0.0340	0.0480	−0.0380	0.0150	0.0470	1		
(14) Services	0.551 ***	0.199 ***	0.316 ***	0.326 ***	0.307 ***	0.376 ***	0.393 ***	−0.099 ***	−0.068 **	0.155 ***	0.291 ***	0.207 ***	0.058 *	1	
(15) Invisits	0.813 ***	0.474 ***	0.449 ***	0.359 ***	0.331 ***	0.352 ***	0.364 ***	−0.121 ***	−0.105 ***	0.420 ***	0.440 ***	0.248 ***	−0.054 *	0.454 ***	1

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5. Regression analysis.

	(1)	(2)	(3)	(4)	(5)
	Patient Selection	**Patient Selection**	**Patient Selection**	**Patient Selection**	**Patient Selection**
HospitalRank	0.004	0.027	0.039	0.039	0.054
	(0.054)	(0.437)	(0.665)	(0.672)	(0.931)
HospitalTitle	−0.31 ***	−0.289 ***	−0.22 **	−0.226 ***	−0.22 ***
	(−3.298)	(−3.25)	(−2.542)	(−2.644)	(−2.585)
DoctorTitle	−0.253 ***	−0.105 *	−0.098 *	−0.097 *	−0.095 *
	(−4.752)	(−1.915)	(−1.844)	(−1.849)	(−1.816)
Hot	0.646 ***	0.589 ***	0.414 **	0.317 *	0.352 **
	(3.525)	(3.487)	(2.495)	(1.935)	(2.15)

Table 5. *Cont.*

	(1)	(2)	(3)	(4)	(5)
	Patient Selection	**Patient Selection**	**Patient Selection**	**Patient Selection**	**Patient Selection**
Honor	−0.293 **	−0.276 **	−0.245 **	−0.203 *	−0.196 *
	(−2.43)	(−2.485)	(−2.27)	(−1.91)	(−1.85)
Reply	0.004	−0.026	−0.033	−0.019	−0.013
	(0.074)	(−0.549)	(−0.714)	(−0.423)	(−0.289)
Services	0.539 ***	0.416 ***	0.38 ***	0.341 ***	0.335 ***
	(11.665)	(9.548)	(8.908)	(8.037)	(7.832)
Invisits	0.885 ***	0.795 ***	0.766 ***	0.766 ***	0.764 ***
	(33.39)	(29.426)	(29.007)	(29.583)	(29.473)
Information quantity		−0.142 ***	−0.082 *	−0.12 ***	−0.113 ***
		(−3.192)	(−1.782)	(−2.824)	(−2.662)
Emotion expression		0.423 ***	0.414 ***	0.622 ***	0.398 ***
		(4.156)	(4.191)	(5.063)	(4.1)
Profile photo		1.098 ***	0.944 ***	0.882 ***	1.026 ***
		(11.966)	(10.39)	(9.806)	(10.144)
ExpertisePre			5.079 ***		
			(4.346)		
Information quantity * ExpertisePre			−0.632 **		
			(−2.277)		
AttitudePre				1.837 ***	
				(9.729)	
Emotion expression * AttitudePre				−0.882 ***	
				(−2.99)	
CredibilityPre					1.237 ***
					(8.292)
Profile photo * CredibilityPre					−0.488 ***
					(−2.663)
_cons	−7.968 ***	−7.211 ***	−6.692 ***	−6.329 ***	−6.535 ***
	(−12.221)	(−11.872)	(−11.241)	(−10.782)	(−11.126)
Observations	1036	1036	1036	1036	1036
R-squared	0.715	0.759	0.774	0.782	0.782

t-values are in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Model (2) adds the independent variables *Information quantity*, *emotion* expression and *profile photo* to model (1). The doctors' information quantity on display negatively affects patient selection ($\beta = -0.142$, $p < 0.01$), with an increase of 0.044 in R-squared, indicating enhanced explanatory power of the model. This result supports H1, suggesting that a blending-in information strategy will enhance the probability of patient selection. It is inappropriate for the doctor to describe his or her expertise through too much professional text, even though it is an important indicator for patients to gauge the doctor's medical skills and ethics [73]. Doctors should have a focused introduction of themselves, focusing on patients' concerns, so that patients can quickly understand the doctor's professional expertise. Too much text description may increase patients' cognitive burden, affecting the extraction of key information. This is not conducive to making a positive impression, thus hindering patient selection. These results are consistent with prior findings [34], with the exception of the result for informational support [56]. Doctors' emotional expression on display positively influences patient selection ($\beta = 0.423$, $p < 0.01$), supporting H2, suggesting that positive affective strategies will enhance the probability of patient selection. A doctor's positive emotional expression is more likely to attract patients. It is common for patients to feel nervous and depressed when they visit the doctor. Positive doctor's messages can be effective at calming patients' emotions and make them feel warm. The doctor can also leave patients with a positive impression by being kind and friendly, thereby increasing the number of patients who choose the doctor. This result is consistent with existing studies [10]. Doctors' profile photo use positively affects patient selection ($\beta = 1.098$, $p < 0.01$), which supports H3, indicating that a positive image strategy will enhance the probability of patient selection. Online consultation breaks the time and space limitations between doctors and patients but reduces doctor–patient trust. A patient's main concern when choosing an internet doctor is their credibility. Doctors who upload real avatars contribute significantly to the construction of their identity and image in the eyes of patients. This authenticity and credibility enhance patients' trust, ultimately attracting more patients to opt for their online consultation services. This is consistent with the findings in the context of online shopping. The photos of merchants on website platforms can influence buyers' perception of their credibility, thereby affecting the establishment of trust [14].

Model (3) adds patient impression and the interaction of patient impression and information quantity of doctor introduction to model (2), resulting in an increase of 0.015 in R-squared, indicating an enhanced explanatory power of the model. Patient impressions of doctors' expertise preference positively moderate the negative effect of information quantity on patient selection ($\beta = -0.632$, $p < 0.05$), as shown in Figure 3, which supports H4a. It suggests that patients' high evaluation of a doctor's expertise magnifies the negative effects of information quantity on their selection. The doctor's introduction primarily serves to assist patients in comprehending the doctor's professional information. On the other hand, the assessment of a doctor's expertise based on patients' impressions complements the information in the doctor's introduction, and these two elements serve as alternatives. Patient impressions represent the collective, more profound perceptual judgments made by other patients regarding the doctor, adding a persuasive dimension to the evaluation. Therefore, when the doctor receives high expertise ratings from other patients, detailed introductions become unnecessary for patients to grasp the doctor's specialization. In this scenario, if the doctor offers an extensive introduction, the surplus information will more significantly diminish the patient's impression of the doctor, ultimately proving counterproductive to the patient's decision-making.

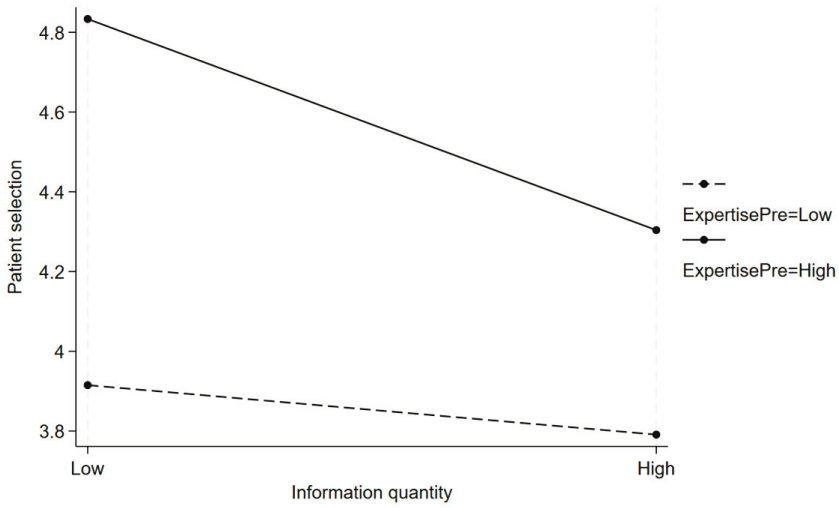


Figure 3. Moderating effect of Expertise Preference in patient impression on information quantity.

Model (4) introduces patient impressions and the interaction of patient impressions with emotional expression to model (2), resulting in an increase of 0.023 in R-squared, demonstrating an improvement in the model's explanatory strength. Patient impressions of doctors' attitudes preference negatively modulate the positive effects of emotional expression on patient selection ($\beta = -0.882$, $p < 0.01$), as shown in Figure 4, supporting H4b. This implies that the favorable impact of positive emotions on patient selection decreases when the patient's evaluation of the doctor's attitude is already favorable. Patients' impressions of a doctor's demeanor are shaped by both the doctor's self-disclosure (doctor's message) and other patients' reviews (patient impressions), both of which can bolster patients' perceptions of doctors and positively affect their selection. However, a substitution effect exists between these factors. When other patients reported that a doctor's attitude was superior, the significance of positive emotional expression conveyed in the doctor's message diminished.

Figure 4. Moderating effect of patient impressions Attitude Preference in patient impression on doctor's emotion expression.

Model (5) adds patient impression and the interaction of patient impression and profile photo to model (2), resulting in an increase of 0.023 in R-squared. Patient impressions of higher credibility negatively moderate the positive effect of profile photo on patient selection ($\beta = -0.488$, $p < 0.01$), as shown in Figure 5, supporting H4c. This implies that the positive impact of a doctor's profile photo on patient selection decreases when the patient

holds a high evaluation of the doctor's credibility. Regarding the doctor's profile photo, the disclosure of the doctor's self-information and patient impressions are substitutable. This result aligns with findings from previous studies, where patient reviews in doctor performance reports can reduce participants' attention to standardized measures of patient experiences and clinical process measures, increasing the likelihood of choosing doctors who perform poorly on those metrics [74]. Our study extends this conclusion, suggesting that it is not only standardized measures but also the impact of emotional expression and profile photos on patient selection which are influenced by patient reviews.

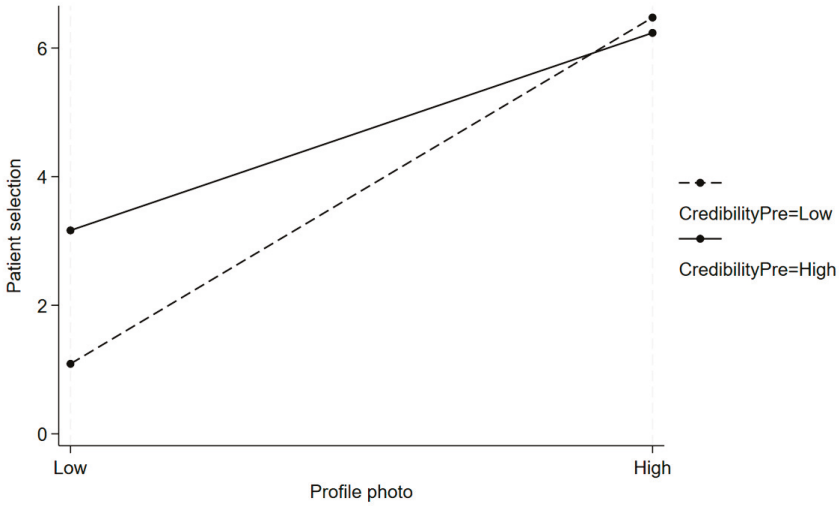


Figure 5. Moderating effect of patient impression Credibility Preference in patient impression on doctors' profile photo.

3.4. Robustness Check

Building upon the original method, we reevaluated the calculation of both the information quantity in the doctor's introduction and the doctor's emotional expression. The information quantity was recalibrated using TextMind. To clarify, the information quantity, which refers to the count of substantive words, is derived by subtracting the number of function words from the total word count. The regression results, presented in Table 6 for models (2) and (3), reveal that the information quantity exerts a negative influence on patient selection, and patient impression effectively moderates the association between the information quantity and patient selection. These findings align with the initial observations.

Emotions are quantified through TextMind, where the emotional polarity of sentiments is computed. The doctor's emotional state is determined by subtracting the number of negative words from the number of positive words. The regression outcomes are presented in Table 7 for models (2) and (4), and indicate that doctors' emotional expression has a positive impact on patient selection, and patient impressions negatively moderate the influence of doctor emotions on patient selection. These conclusions align with the initial findings.

Apart from employing patient impressions as a measure of patient feedback, patient ratings can also serve this purpose. In this section, favorable doctor ratings are utilized in place of patient impressions to assess their moderating impact on information quantity, emotion expression, and profile photo. The correlation coefficients between a doctor's favorable rating and the other variables are below 0.6, and there is no problem of multicollinearity. As can be seen in Table 8, all the results are the same as the above regression results except for model (3). In model (3), the positive moderating effect of patient ratings on the relationship between the information quantity and patient selection is not

significant. One potential explanation is that, when patient impressions were employed as moderating variables, they were categorized based on various impression contents. In contrast, a doctor's favorable ratings were singular determined values, lacking a specific correspondence between the three moderating variables and the precise aspect of concern of the independent variable. This might have resulted in less accurate outcomes compared to those of the primary model. Regarding the overall robustness of the results, we maintain that the findings of this study are robust.

Table 6. Alternative measure of information quantity (Robustness Check).

	(1)	(2)	(3)	(4)	(5)
	Patient Selection	**Patient Selection**	**Patient Selection**	**Patient Selection**	**Patient Selection**
HospitalRank	−0.026	−0.002	0.017	0.017	0.032
	(−0.395)	(−0.025)	(0.291)	(0.289)	(0.555)
HospitalTitle	0.341 ***	0.31 ***	0.248 ***	0.239 ***	0.24 ***
	(4.122)	(3.977)	(3.271)	(3.186)	(3.213)
DoctorTitle	−0.259 ***	−0.109 **	−0.099 *	−0.101 *	−0.098 *
	(−4.872)	(−2)	(−1.88)	(−1.947)	(−1.886)
Hot	0.65 ***	0.593 ***	0.42 **	0.323 **	0.358 **
	(3.558)	(3.517)	(2.538)	(1.974)	(2.191)
Honor	−0.294 **	−0.279 **	−0.247 **	−0.206 *	−0.199 *
	(−2.448)	(−2.519)	(−2.298)	(−1.944)	(−1.883)
Reply	0.006	−0.024	−0.032	−0.018	−0.011
	(0.118)	(−0.503)	(−0.677)	(−0.386)	(−0.251)
Services	0.541 ***	0.418 ***	0.382 ***	0.344 ***	0.337 ***
	(11.735)	(9.616)	(8.975)	(8.102)	(7.904)
Invisits	0.883 ***	0.795 ***	0.767 ***	0.766 ***	0.765 ***
	(33.399)	(29.452)	(29.034)	(29.575)	(29.485)
Information quantity		−0.143 ***	−0.085 *	−0.118 ***	−0.114 ***
		(−3.353)	(−1.931)	(−2.898)	(−2.809)
Emotion expression		0.425 ***	0.417 ***	0.619 ***	0.4 ***
		(4.188)	(4.236)	(5.06)	(4.131)
Profile photo		1.088 ***	0.937 ***	0.876 ***	1.022 ***
		(11.892)	(10.331)	(9.765)	(10.121)
ExppertisePre			5.341 ***		
			(4.287)		
Information quantity * ExppertisePre			−0.653 **		
			(−2.369)		
AttitudePre				1.814 ***	
				(9.606)	
Emotion expression * AttitudePre				−0.868 ***	
				(−2.952)	
CredibilityPre					1.23 ***
					(8.262)
Profile photo * CredibilityPre					−0.494 ***
					(−2.701)
_cons	−8.504 ***	−7.674 ***	−7.078 ***	−6.699 ***	−6.903 ***
	(−13.018)	(−12.637)	(−11.847)	(−11.349)	(−11.699)
Observations	1036	1036	1036	1036	1036
R-squared	0.717	0.76	0.775	0.782	0.782

t-values are in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7. Alternative measure of emotion expression (Robustness Check).

	(1)	(2)	(3)	(4)	(5)
	Patient Selection	**Patient Selection**	**Patient Selection**	**Patient Selection**	**Patient Selection**
HospitalRank	−0.026	−0.002	0.018	0.017	0.032
	(−0.395)	(−0.029)	(0.302)	(0.291)	(0.545)
HospitalTitle	0.341 ***	0.289 ***	0.227 ***	0.222 ***	0.22 ***
	(4.122)	(3.703)	(2.98)	(2.944)	(2.935)
DoctorTitle	−0.259 ***	−0.118 **	−0.11 **	−0.102 *	−0.107 **
	(−4.872)	(−2.152)	(−2.062)	(−1.952)	(−2.038)
Hot	0.65 ***	0.615 ***	0.434 ***	0.322 *	0.376 **
	(3.558)	(3.623)	(2.604)	(1.953)	(2.282)
Honor	−0.294 **	−0.28 **	−0.249 **	−0.217 **	−0.201 *
	(−2.448)	(−2.52)	(−2.299)	(−2.042)	(−1.885)

Table 7. *Cont.*

	(1)	(2)	(3)	(4)	(5)
	Patient Selection	**Patient Selection**	**Patient Selection**	**Patient Selection**	**Patient Selection**
Reply	0.006	−0.024	−0.028	−0.015	−0.01
	(0.118)	(−0.494)	(−0.606)	(−0.317)	(−0.226)
Services	0.541 ***	0.429 ***	0.392 ***	0.347 ***	0.348 ***
	(11.735)	(9.842)	(9.178)	(8.169)	(8.118)
Invisits	0.883 ***	0.816 ***	0.788 ***	0.786 ***	0.785 ***
	(33.399)	(31.047)	(30.641)	(31.1)	(31.056)
Information quantity		−0.132 ***	−0.078 *	−0.115 ***	−0.103 **
		(−2.982)	(−1.687)	(−2.717)	(−2.444)
Emotion expression		2.824 ***	2.444 **	4.803 ***	2.459 **
		(2.734)	(2.432)	(3.521)	(2.493)
Profile photo		1.131 ***	0.983 ***	0.928 ***	1.062 ***
		(12.409)	(10.871)	(10.387)	(10.532)
ExpertisePre			4.703 ***		
			(4.005)		
Information quantity * ExpertisePre			−0.545 *		
			(−1.953)		
AttitudePre				1.663 ***	
				(9.905)	
Emotion expression * AttitudePre				−7.943 **	
				(−2.572)	
CredibilityPre					1.23 ***
					(8.209)
Profile photo * CredibilityPre					−0.488 ***
					(−2.653)
_cons	−8.504 ***	−7.961 ***	−7.325 ***	−6.879 ***	−7.164 ***
	(−13.018)	(−13.123)	(−12.256)	(−11.655)	(−12.142)
Observations	1036	1036	1036	1036	1036
R-squared	0.717	0.758	0.772	0.779	0.78

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8. Alternative measure of patient impressions (Robustness Check).

	(1)	(2)	(3)	(4)	(5)	(6)
	Patient Selection	**Patient Selection**	**Patient Selection**	**Patient Selection**	**Patient Selection**	**Patient Selection**
Hospital_Rank	0.057	0.066	0.061	0.064	0.059	0.058
	(0.935)	(1.155)	(1.07)	(1.117)	(1.053)	(1.03)
Hospital_Title	−0.29 ***	−0.267 ***	−0.27 ***	−0.273 ***	−0.278 ***	−0.281 ***
	(−3.324)	(−3.19)	(−3.232)	(−3.285)	(−3.406)	(−3.449)
Doctor_Title	−0.185 ***	−0.079	−0.084	−0.082	−0.1 **	−0.1 **
	(−3.726)	(−1.518)	(−1.622)	(−1.592)	(−1.97)	(−1.988)
hot	−0.141	−0.065	−0.045	0.004	0.084	0.116
	(−0.78)	(−0.386)	(−0.265)	(0.025)	(0.508)	(0.7)
honor	−0.216 *	−0.212 **	−0.209 **	−0.181 *	−0.173 *	−0.157
	(−1.933)	(−2.025)	(−1.998)	(−1.739)	(−1.69)	(−1.537)
reply	−0.042	−0.06	−0.066	−0.048	−0.046	−0.041
	(−0.869)	(−1.324)	(−1.456)	(−1.077)	(−1.043)	(−0.915)
Services	0.264 ***	0.2 ***	0.208 ***	0.209 ***	0.22 ***	0.224 ***
	(5.519)	(4.446)	(4.598)	(4.659)	(4.998)	(5.073)
Invisits	0.816 ***	0.745 ***	0.744 ***	0.74 ***	0.747 ***	0.744 ***
	(32.473)	(28.951)	(28.907)	(28.865)	(29.764)	(29.608)
GoodRate	1.821 ***	1.54 ***	2.166 ***	1.867 ***	2.777 ***	2.926 ***
	(13.051)	(11.607)	(5.16)	(11.633)	(13.09)	(6.882)
Information quantity		−0.102 **	−0.075 *	−0.102 **	−0.11 ***	−0.108 **
		(−2.443)	(−1.647)	(−2.458)	(−2.684)	(−2.416)
Emotion expression		0.366 ***	0.378 ***	0.622 ***	0.359 ***	0.511 ***
		(3.816)	(3.931)	(5.219)	(3.842)	(4.321)
Profile photo		0.942 ***	0.934 ***	0.914 ***	1.259 ***	1.222 ***
		(10.789)	(10.692)	(10.488)	(13.195)	(12.576)
Information quantity * GoodRate			−0.157			−0.009
			(−1.572)			(−0.092)
Emotion expression * GoodRate				−0.77 ***		−0.452 **
				(−3.58)		(−2.045)
Profile photo * GoodRate					−1.738 ***	−1.626 ***
					(−7.357)	(−6.696)
_cons	−4.858 ***	−4.701 ***	−4.803 ***	−4.956 ***	−5.31 ***	−5.427 ***
	(−7.481)	(−7.697)	(−7.825)	(−8.105)	(−8.832)	(−8.98)
Observations	1036	1036	1036	1036	1036	1036
R-squared	0.756	0.787	0.788	0.79	0.798	0.799

t-values are in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4. Discussion

4.1. Theoretical Implications

Our research has the following theoretical implications: First, this study focuses on the disclosure strategies employed by doctors on online medical platforms, delving into how these strategies influence patients' selections from the perspective of impression management. Prior research has demonstrated that the amount of information presented on a doctor's homepage has an adverse impact on patient decision-making, whereas positive emotions displayed by the doctor tend to attract patients [10]. This paper provides a more detailed differentiation among various components of a doctor's homepage, considering three distinct aspects of patients' attention: information management, affective management, and image management. The study finds that implementing a blending-in information strategy, a positive affective strategy, and a positive image strategy promotes patient selection. Recognizing the important role of image elements in doctor disclosure, our research indicates that patients are more likely to select doctors when they use real images. This also suggests that doctors need to focus on the integration of multimodal information and enhance the comprehensive use of images, text, and other disclosure methods [35]. Moreover, our study indicates that positive strategies and blending-in strategies are not mutually exclusive. By flexibly applying these two strategies across the dimensions of information, affect, and image management, doctors can achieve positive results in attracting patients. The study enhances our understanding of information disclosure on online medical platforms, offering fresh perspectives on the relationship between doctors' homepage self-presentation and patient selection. It also guides further research on homepage design and related topics on online platforms.

Second, this study examines the moderating effect of patient impressions on doctors' self-disclosure from the perspective of information integration and categorizes patient impressions creatively to analyze the moderating effects of different patient impression categories on different independent variables separately. The results suggest that both doctors' self-presentation and other patient reviews are important factors influencing patient service selection. This study reveals in more detail the influence mechanisms between doctor self-presentation, patient reviews and subsequent patient selection. The findings provide new ideas for subsequent research on patient feedback and the integration of multiple aspects of information.

Third, our study expands the application scenarios of impression management theory. In past studies, impression management has been widely applied to online social media [57,75], corporate media platforms [16], and so on. This study applies it to the impression management of doctors on online medical platforms, contending that doctors' homepage impression management significantly impacts patients' selection. Extending the impression management theory to the online medical domain enhances the breadth of impression management research and offers fresh insights for the future application of this theory.

4.2. Practical Implications

The research findings presented in this study also have practical implications. First, for doctors, our findings offer practical guidance for effective homepage management and the cultivation of high-quality professional profiles, helping ordinary doctors effectively attract patients in the competitive online medical environment. (1) For doctors aiming to boost patient selection and expand their influence, it is advisable to keep the text information in their introductions concise, emphasizing their professional expertise. Particularly for doctors with a highly rated level of expertise according to patients' impressions, they should be mindful of the length of their introductions. Excessive information in doctors'

introductions can inadvertently create a barrier between doctors and patients, potentially hindering patients' selection. (2) Doctors should enhance their expression of positive emotions on the online medical platform by demonstrating their approachability. They should also maintain a positive attitude towards medicine and effectively reassure patients to alleviate their concerns. (3) Doctors should also focus on the image features of multimodal information, improve their image information as much as possible, and increase credibility by providing a realistic image.

Second, for online medical platforms, designing doctors' homepages in a thoughtful manner is crucial. It involves a focus on the integration of multimedia elements like images and videos, as well as providing doctors with greater autonomy and sharing capabilities. At the same time, more feedback channels should be given to the patients who have already attended a consultation. Theme clustering and labeling can provide patients who have not been consulted with the doctor's impression from different dimensions. They can also provide patients with more opinions for reference, so that they can make a more suitable selection for themselves.

5. Conclusions

The behaviors of both doctors and patients on online medical platforms are garnering growing attention within the realm of internet medical research. It is important for improving the effective operation of online platforms. To address the two questions posed at the beginning of the paper, this study investigates the influence of doctors' homepage impression management strategies on patient selection and delves into the interaction between a doctor's strategies for information, affective, image management, and the patient impressions based on data from a psychiatric department. The following conclusions are drawn: (1) The blending-in information strategy enhances patient selection, as a greater quantity of information in a doctor's introduction has a detrimental impact on patient selection. The positive affective strategy has a beneficial impact on patient selection, as the positive emotion conveyed in a doctor's message effectively influences patient decisions. The positive image strategy contributes to enhancing patient selection, as the genuine image of the doctor positively influences patient selection. (2) Patient impressions related to doctors' expertise amplify the negative effect of information quantity on patient selection, while patient impressions related to doctors' attitude and credibility separately negatively moderate the relationship between emotional expression and patient selection, and between real-photo avatars and patient selection. Doctors can positively adjust their disclosures based on patient impressions and pay attention to the integrating role of multifaceted information.

The current study has the following limitations. First, this study only explores the impact of doctors' impression management on patient selection. Future research could further investigate whether and how doctors' impression management affects patient satisfaction. Although patient satisfaction is primarily influenced by doctor–patient interactions, the initial impression of a doctor may also impact how patients evaluate them. Furthermore, this study mainly focuses on the interaction between doctor information disclosure, impression management, and patient feedback in a static context. However, it is important to note that doctors may dynamically adjust their impression management strategies accordingly based on patient feedback or patient selection, which is a process that the current data cannot capture. Therefore, future research could further capture the dynamic changes in doctors' impression management strategies, patient impressions, and patient selection, in order to explore the dynamic impact of impression management strategies.

Author Contributions: Conceptualization, X.L.; Methodology, X.L. and S.Z.; Formal Analysis, S.Z. and X.C.; Data Curation, X.C.; Writing—Original Draft Preparation, X.L., S.Z., X.C. and Y.C.; Writing—Review & Editing, X.L., X.C., J.L. and Y.C.; Funding Acquisition, X.L. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant Nos. 71971082, 72271093), the Science and Technology Innovation Plan of Shanghai Science and Technology Commission (Grant Nos. 22692110200, 19692106700) and Humanity and Social Science Youth foundation of Ministry of Education of China (Grant No. 18YJC630068).

Institutional Review Board Statement: Due to the nature of the study, the Ethics Committee approval wasn't required.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Young, C. Community Management That Works: How to Build and Sustain a Thriving Online Health Community. *J. Med. Internet Res.* **2013**, *15*, e119. [CrossRef] [PubMed]
2. Guo, S.S.; Guo, X.T.; Fang, Y.L.; Vogel, D. How Doctors Gain Social and Economic Returns in Online Health-Care Communities: A Professional Capital Perspective. *J. Manag. Inf. Syst.* **2017**, *34*, 487–519. [CrossRef]
3. Fan, H.; Lederman, R.; Smith, S.; Chang, S. How Trust Is Formed in Online Health Communities: A Process Perspective. *Commun. Assoc. Inf. Syst.* **2014**, *34*, 531–560. [CrossRef]
4. Wu, H.; Deng, Z.; Wang, B.; Wang, H. How Online Health Community Participation Affects Physicians' Performance in Hospitals: Empirical Evidence from China. *Inf. Manag.* **2021**, *58*, 103443. [CrossRef]
5. Fang, J.; Wen, L.; Ren, H.; Wen, C. The effect of technical and functional quality on online physician selection: Moderation effect of competition intensity. *Inf. Process. Manag.* **2022**, *59*, 102969. [CrossRef]
6. Introduction to Haodf.com. Available online: https://www.haodf.com/info/aboutus.php (accessed on 6 January 2025).
7. Li, J.; Zhang, Y.; Ma, L.; Liu, X. The impact of the internet on health consultation market concentration: An econometric analysis of secondary data. *J. Med. Internet Res.* **2016**, *18*, e276. [CrossRef] [PubMed]
8. Goffman, E. *The Presentation of Self in Everyday Life*; Doubleday: Garden City, NY, USA, 1959; ISBN 978-03-8509-402-3.
9. Vitak, J. Balancing Privacy Concerns and Impression Management Strategies on Facebook. In Proceedings of the Symposium on Usable Privacy and Security (SOUPS), Ottawa, ON, Canada, 22–24 July 2015; USENIX: Berkeley, CA, USA, 2015; pp. 22–24.
10. Ouyang, P.; Wang, J.-J.; Chang, A.-C. Patients Need Emotional Support: Managing Physician Disclosure Information to Attract More Patients. *Int. J. Med. Inf.* **2022**, *158*, 104674. [CrossRef]
11. Rui, J.R.; Stefanone, M.A. Strategic Image Management Online: Self-Presentation, Self-Esteem and Social Network Perspectives. *Inf. Commun. Soc.* **2013**, *16*, 1286–1305. [CrossRef]
12. Yang, J.; Li, Y.; Calic, G.; Shevchenko, A. How Multimedia Shape Crowdfunding Outcomes: The Overshadowing Effect of Images and Videos on Text in Campaign Information. *J. Bus. Res.* **2020**, *117*, 6–18. [CrossRef]
13. Hernández-Ortega, B. Don't believe strangers: Online consumer reviews and the role of social psychological distance. *Inf. Manag.* **2018**, *55*, 31–50. [CrossRef]
14. Leong, L.-Y.; Hew, T.-S.; Ooi, K.-B.; Chong, A.Y.L.; Lee, V.-H. Understanding trust in ms-commerce: The roles of reported experience, linguistic style, profile photo, emotional, and cognitive trust. *Inf. Manag.* **2021**, *58*, 103416. [CrossRef]
15. Lynch, J.W.; Rodell, J.B. Blend in or Stand Out? Interpersonal Outcomes of Managing Concealable Stigmas at Work. *J. Appl. Psychol.* **2018**, *103*, 1307–1323. [CrossRef]
16. Sun, Y.; Fang, S.; Zhang, Z.J. Impression Management Strategies on Enterprise Social Media Platforms: An Affordance Perspective. *Int. J. Inf. Manag.* **2021**, *60*, 102359. [CrossRef]
17. Liu, X.; Zhou, S.; Chi, X. How Do Team-Level and Individual-Level Linguistic Styles Affect Patients' Emotional Well-Being—Evidence from Online Doctor Teams. *Int. J. Environ. Res. Public Health* **2023**, *20*, 1915. [CrossRef]
18. de Vaate, A.J.D.B.; Veldhuis, J.; Alleva, J.M.; Konijn, E.A.; van Hugten, C.H.M. Show Your Best Self(ie): An Exploratory Study on Selfie-Related Motivations and Behavior in Emerging Adulthood. *Telemat. Inf.* **2018**, *35*, 1392–1407. [CrossRef]

19. Liu, X.; Chen, M.; Li, J.; Ma, L. How to Manage Diversity and Enhance Team Performance: Evidence from Online Doctor Teams in China. *Int. J. Environ. Res. Public Health* **2019**, *17*, 48. [CrossRef]
20. Meng, F.; Zhang, X.; Liu, L.; Ren, C. Converting Readers to Patients? From Free to Paid Knowledge-Sharing in Online Health Communities. *Inf. Process. Manag.* **2021**, *58*, 102490. [CrossRef]
21. Guo, L.; Jin, B.; Yao, C.; Yang, H.; Huang, D.; Wang, F. Which doctor to trust: A recommender system for identifying the right doctors. *J. Med. Internet Res.* **2016**, *18*, e6015. [CrossRef] [PubMed]
22. Wang, Z.; Liang, L.; Liu, X.; Liao, M. Research on the influencing factors and the differences between the initial trust and continuous trust of online health community users. *Neural Comput. Appl.* **2024**, *36*, 9849–9874. [CrossRef]
23. Li, J.; Tang, J.; Jiang, L.; Yen, D.C.; Liu, X. Economic success of physicians in the online consultation market: A signaling theory perspective. *Int. J. Electron. Commer.* **2019**, *23*, 244–271. [CrossRef]
24. Anderson, N.H. Foundations of Information Integration Theory. *Am. J. Psychol.* **1981**, *95*, 708–711.
25. Mudrik, L.; Faivre, N.; Koch, C. Information Integration without Awareness. *Trends Cogn. Sci.* **2014**, *18*, 488–496. [CrossRef]
26. Anderson, N.H. Integration Theory and Attitude Change. *Psychol. Rev.* **1971**, *78*, 171–206. [CrossRef]
27. Wang, P.; Fan, W.L.; Li, Q. The Alleviation Effect of Knowledge Integration on Anxiety in Parental Vaccine Decision-Making Supported by Online Communities. *J. Knowl. Manag.* **2021**, *25*, 1575–1593. [CrossRef]
28. Wong, C.W.V.; Lai, K.H.; Cheng, T.C.E.; Lun, Y.H.V. The Role of IT-Enabled Collaborative Decision Making in Inter-organizational Information Integration to Improve Customer Service Performance. *Int. J. Prod. Econ.* **2015**, *159*, 56–65. [CrossRef]
29. Sanders, R.; Araujo, T.B.; Vliegenthart, R.; van Eenbergen, M.C.; van Weert, J.C.M.; Linn, A.J. Patients' Convergence of Mass and Interpersonal Communication on an Online Forum: Hybrid Methods Analysis. *J. Med. Internet Res.* **2020**, *22*, e18303. [CrossRef]
30. Chen, S.; Guo, X.; Wu, T.; Ju, X. Exploring the Online Doctor-Patient Interaction on Patient Satisfaction Based on Text Mining and Empirical Analysis. *Inf. Proc. Manag.* **2020**, *57*, 102253. [CrossRef]
31. Xia, A.; Perera, S.C.; Ahmed, M.U.; Tang, J.; Wang, J.J. Voice or text? The role of physician media choice on patient experience in online medical communities. *Decis. Sci.* **2024**, *55*, 620–638. [CrossRef]
32. Jiang, S.; Liu, X.; Chi, X. Effect of Writing Style on Social Support in Online Health Communities: A Theoretical Linguistic Analysis Framework. *Inf. Manag.* **2022**, *59*, 103683. [CrossRef]
33. Liu, S.; Zhang, M.; Gao, B.; Jiang, G. Physician Voice Characteristics and Patient Satisfaction in Online Health Consultation. *Inf. Manag.* **2020**, *57*, 103233. [CrossRef]
34. Wu, H.; Deng, Z.; Evans, R. Building Patients' Trust in Psychologists in Online Mental Health Communities. *Data Sci. Manag.* **2022**, *5*, 21–27. [CrossRef]
35. Zhou, Y.; Zhu, L.; Wu, C.-N.; Huang, S.; Wang, Q. Do the Rich Grow Richer? An Empirical Analysis of the Matthew Effect in an Online Healthcare Community. *Electron. Commer. Res. Appl.* **2022**, *52*, 101125. [CrossRef]
36. Carlson, D.S.; Kacmar, K.M.; Thompson, M.J.; Andrews, M.C. Looking Good and Doing Good: Family to Work Spillover through Impression Management. *J. Manag. Psychol.* **2019**, *34*, 31–45. [CrossRef]
37. Bolino, M.C. Citizenship and Impression Management: Good Soldiers or Good Actors? *Acad. Manag. Rev.* **1999**, *24*, 82–98. [CrossRef]
38. Benthaus, J.; Risius, M.; Beck, R. Social media management strategies for organizational impression management and their effect on public perception. *J. Strateg. Inf. Syst.* **2016**, *25*, 127–139. [CrossRef]
39. Lee, G.; Cho, S.Y.; Arthurs, J.; Lee, E.K. Celebrity CEO, identity threat, and impression management: Impact of celebrity status on corporate social responsibility. *J. Bus. Res.* **2020**, *111*, 69–84. [CrossRef]
40. Schniederjans, D.; Cao, E.S.; Schniederjans, M. Enhancing financial performance with social media: An impression management perspective. *Decis. Support Syst.* **2013**, *55*, 911–918. [CrossRef]
41. Yang, J.H.; Liu, S. Accounting narratives and impression management on social media. *Acc. Bus. Res.* **2017**, *47*, 673–694. [CrossRef]
42. Sezer, O.; Gino, F.; Norton, M.I. Humblebragging: A distinct—And ineffective—Self-presentation strategy. *J. Pers. Soc. Psychol.* **2018**, *114*, 52. [CrossRef] [PubMed]
43. Roulin, N.; Bangerter, A.; Levashina, J. Honest and deceptive impression management in the employment interview: Can it be detected and how does it impact evaluations? *Pers. Psychol.* **2015**, *68*, 395–444. [CrossRef]
44. Roulin, N.; Bangerter, A.; Levashina, J. Interviewers' perceptions of impression management in employment interviews. *J. Manag. Psychol.* **2014**, *29*, 141–163. [CrossRef]
45. Chen, J.; Ping, J.W.; Xu, Y.C.; Tan, B.C.Y. Information Privacy Concern About Peer Disclosure in Online Social Networks. *IEEE Trans. Eng. Manag.* **2015**, *62*, 311–324. [CrossRef]
46. McDonnell, M.-H.; King, B. Keeping Up Appearances: Reputational Threat and Impression Management after Social Movement Boycotts. *Adm. Sci. Q.* **2013**, *58*, 387–419. [CrossRef]

47. Chen, J.E.; Ping, J.W.; Xu, Y.C.; Tan, B.C.Y. Am I Afraid of My Peers? Understanding the Antecedents of Information Privacy Concerns in the Online Social Context. In Proceedings of the ICIS. 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems, Shanghai, China, 20–22 November 2009.
48. Bolino, M.C.; Kacmar, K.M.; Turnley, W.H.; Gilstrap, J.B. A multi-level review of impression management motives and behaviors. *J. Manag.* **2008**, *34*, 1080–1109. [CrossRef]
49. Li, X.; Wang, Q.; Yao, X.; Yan, X.; Li, R. How Do Influencers' Impression Management Tactics Affect Purchase Intention in Live Commerce?–Trust Transfer and Gender Differences. *Inf. Manag.* **2024**, *62*, 104094. [CrossRef]
50. Cheng, J.-W.; Chiu, W.-L.; Tzeng, G.-H. Do impression management tactics and/or supervisor–subordinate guanxi matter? *Knowl.-Based Syst.* **2013**, *40*, 123–133. [CrossRef]
51. Ollier-Malaterre, A.; Rothbard, N.P.; Berg, J.M. When Worlds Collide in Cyberspace: How Boundary Work in Online Social Networks Impacts Professional Relationships. *Acad. Manag. Rev.* **2013**, *38*, 645–669. [CrossRef]
52. Rosenberg, J.; Egbert, N. Online Impression Management: Personality Traits and Concerns for Secondary Goals as Predictors of Self-Presentation Tactics on Facebook. *J. Comput. Mediat. Commun.* **2011**, *17*, 1–18. [CrossRef]
53. Fan, J.; Geng, H.; Liu, X.; Wang, J. The effects of online text comments on patients' choices: The mediating roles of comment sentiment and comment content. *Front. Psychol.* **2022**, *13*, 886077. [CrossRef]
54. Chen, L.T.; Baird, A.; Straub, D. A Linguistic Signaling Model of Social Support Exchange in Online Health Communities. *Decis. Support Syst.* **2020**, *130*, 113233. [CrossRef]
55. Bazarova, N.N.; Taft, J.G.; Choi, Y.H.; Cosley, D. Managing Impressions and Relationships on Facebook. *J. Lang. Soc. Psychol.* **2013**, *32*, 121–141. [CrossRef]
56. Bell, R.; Daly, J. The Affinity-Seeking Function of Communication. *Commun. Monogr.* **1984**, *51*, 91–115. [CrossRef]
57. Rubin, R.; Rubin, A.; Martin, M. The Role of Self-Disclosure and Self-Awareness in Affinity-Seeking Competence. *Commun. Res. Rep.* **1993**, *10*, 115–127. [CrossRef]
58. Liu, X.; Jiang, S.; Sun, M.; Chi, X. Examining Patterns of Information Exchange and Social Support in a Web-Based Health Community: Exponential Random Graph Models. *J. Med. Internet Res.* **2020**, *22*, e18062. [CrossRef] [PubMed]
59. Lu, Y.; Pan, T.-t.; Liu, J.; Wu, J. Does Usage of Online Social Media Help Users With Depressed Symptoms Improve Their Mental Health? Empirical Evidence From an Online Depression Community. *Front. Public Health* **2020**, *8*, 2296–2565. [CrossRef]
60. Kienzle, J. The Selfie (and Ussie) as Identity in the Classroom. In *Unplugging the Classroom*; Pixy Ferris, S., Wilder, H., Eds.; Chandos Publishing: Cambridge, MA, USA, 2017; pp. 93–102.
61. KrÃ¤mer, N.; Winter, S. Impression Management 2.0: The Relationship of Self-Esteem, Extraversion, Self-Efficacy, and Self-Presentation Within Social Networking Sites. *J. Media Psychol.* **2008**, *20*, 106. [CrossRef]
62. Scott, G.G. More Than Friends: Popularity on Facebook and its Role in Impression Formation. *J. Comput. Mediat. Commun.* **2014**, *19*, 358–372. [CrossRef]
63. Luo, X.M.; Gu, B.; Zhang, J.; Phang, C.W. Expert Blogs and Consumer Perceptions of Competing Brands. *MIS Q.* **2017**, *41*, 371–395. [CrossRef]
64. Ho-Dac, N.; Carson, S.; Moore, W. The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *J. Mark.* **2013**, *77*, 37–53. [CrossRef]
65. Liu, Y. Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *J. Mark.* **2006**, *70*, 74–89. [CrossRef]
66. Cheng, Y.H.; Ho, H.Y. Social Influence's Impact on Reader Perceptions of Online Reviews. *J. Bus. Res.* **2015**, *68*, 883–887. [CrossRef]
67. Wang, Y.; Wu, H.; Lei, X.; Shen, J.; Feng, Z. The Influence of Doctors' Online Reputation on the Sharing of Outpatient Experiences: Empirical Study. *J. Med. Internet Res.* **2020**, *22*, e16691. [CrossRef] [PubMed]
68. Őri, D.; Szocsics, P.; Molnár, T.; Motlova, L.B.; Kazakova, O.; Mörkl, S.; Wallies, M.; Abdulhakim, M.; Boivin, S.; Bruna, K. Attitudes of psychiatrists towards people with mental illness: A cross-sectional, multicentre study of stigma in 32 European countries. *EClinicalMedicine* **2023**, *66*, 102342. [CrossRef] [PubMed]
69. Thornicroft, G.; Sunkel, C.; Aliev, A.A.; Baker, S.; Brohan, E.; El Chammay, R.; Davies, K.; Demissie, M.; Duncan, J.; Fekadu, W. The Lancet Commission on ending stigma and discrimination in mental health. *Lancet* **2022**, *400*, 1438–1480. [CrossRef] [PubMed]
70. Letzring, T.D.; Wells, S.M.; Funder, D.C. Information quantity and quality affect the realistic accuracy of personality judgment. *J. Pers. Soc. Psychol.* **2006**, *91*, 111. [CrossRef]
71. Liu, X.; Chi, X.; Chen, M.; Sun, W.; Li, J. Spillover Effects of Paid Functions on Physicians' Unpaid Knowledge Activities: Quasi-Experimental Approach. *J. Med. Internet Res.* **2024**, *26*, e58688. [CrossRef] [PubMed]
72. Chen, Q.; Jin, J.; Yan, X. Understanding online review behaviors of patients in online health communities: An expectation-disconfirmation perspective. *Inf. Technol. People* **2022**, *35*, 2441–2469. [CrossRef]
73. Wilson, E.J.; Sherrell, D.L. Source Effects in Communication and Persuasion Research: A Meta-Analysis of Effect Size. *J. Acad. Mark. Sci.* **1993**, *21*, 101. [CrossRef]

74. Kanouse, D.E.; Schlesinger, M.; Shaller, D.; Martino, S.C.; Rybowski, L. How patient comments affect consumers' use of physician performance measures. *Med. Care* **2016**, *54*, 24–31. [CrossRef]
75. Tong, S.T.; Heide, B.V.D.; Langwell, L.; Walther, J.B. Too Much of a Good Thing? The Relationship Between Number of Friends and Interpersonal Impressions on Facebook. *J. Comput. Mediat. Commun.* **2008**, *13*, 531–549. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Is Artificial Intelligence a Game-Changer in Steering E-Business into the Future? Uncovering Latent Topics with Probabilistic Generative Models

Simona-Vasilica Oprea * and Adela Bâra

Department of Economic Informatics and Cybernetics, Bucharest University of Economic Studies, Piaţa Romană, 010374 Bucharest, Romania; bara.adela@ie.ase.ro
* Correspondence: simona.oprea@csie.ase.ro

Abstract: Academic publications from the Web of Science Core Collection on "e-business" and "artificial intelligence" (AI) are investigated to reveal the role of AI, extract latent themes and identify potential research topics. The proposed methodology includes relevant graphical representations (trends, co-occurrence networks, Sankey diagrams), sentiment analyses and latent topics identification. A renewed interest in these publications is evident post-2018, with a sharp increase in publications around 2020 that can be attributed to the COVID-19 pandemic. Chinese institutions dominate the collaboration network in e-business and AI. Keywords such as "business transformation", "business value" and "e-business strategy" are prominent, contributing significantly to areas like "Operations Research & Management Science". Additionally, the keyword "e-agribusiness" recently appears connected to "Environmental Sciences & Ecology", indicating the application of e-business principles in sustainable practices. Although three sentiment analysis methods broadly agree on key trends, such as the rise in positive sentiment over time and the dominance of neutral sentiment, they differ in detail and focus. Custom analysis reveals more pronounced fluctuations, whereas VADER and TextBlob present steadier and more subdued patterns. Four well-balanced topics are identified with a coherence score of 0.66 using Latent Dirichlet Allocation, which is a probabilistic generative model designed to uncover hidden topics in large text corpora: Topic 1 (29.8%) highlights data-driven decision-making in e-business, focusing on AI, information sharing and technology-enabled business processes. Topic 2 (28.1%) explores AI and Machine Learning (ML) in web-based business, emphasizing customer service, innovation and workflow optimization. Topic 3 (23.6%) focuses on analytical methods for decision-making, using data modeling to enhance strategies, processes and sustainability. Topic 4 (18.5%) examines the semantic web, leveraging ontologies and knowledge systems to improve intelligent systems and web platforms. New pathways such as voice assistance, augmented reality and dynamic marketplaces could further enhance e-business strategies.

Keywords: e-business; AI; latent themes; sentiment analysis; future pathways; relevant trends

1. Introduction

1.1. General Context of AI Implementation in E-Business

E-business refers to the use of internet-based technologies and digital platforms to conduct business operations. It encompasses a broad range of activities, including online sales (e-commerce), supply chain management, digital marketing, Customer Relationship

Management (CRM) [1,2], etc. AI has emerged as a game-changer in the e-business landscape by enabling automation, personalization, data-driven decision-making and enhanced customer experiences [3,4]. One of the most significant contributions of AI to e-business lies in the customer experience. AI enables businesses to analyze customer behavior, preferences and purchasing history to deliver highly personalized recommendations. Chatbots and virtual assistants, powered by AI, are increasingly used to provide real-time support, guiding users through purchasing decisions and addressing their queries [5–7]. Additionally, dynamic pricing models leverage AI to adjust prices based on demand, customer segmentation and competitor strategies [8]. AI also revolutionizes supply chain and inventory management in e-business. By utilizing predictive analytics, businesses optimize logistics, ensuring timely deliveries and cost efficiency. ML models help predict inventory needs based on historical data, market trends and seasonal fluctuations, reducing over- or understocking risks [9,10]. Furthermore, in terms of security, AI is an important tool for fraud detection and cybersecurity. Advanced algorithms continuously monitor transactions to detect anomalies, safeguarding both businesses and customers from financial losses [11].

AI-powered tools also facilitate smarter data analytics and decision-making. These systems process vast amounts of data in real-time, providing actionable insights that enable businesses to make informed strategic decisions. Predictive analytics helps forecast trends, customer behavior and market dynamics [12,13]. Marketing and customer engagement have also benefited greatly from AI integration. It enhances digital marketing strategies through targeted advertisements, personalized email campaigns and optimized content delivery. With Natural Language Processing (NLP), businesses analyze social media sentiment and customer reviews, allowing them to adapt their strategies based on customer feedback [14].

Routine tasks within e-business are increasingly automated using robotic process automation powered by AI. These systems handle repetitive activities such as data entry, order processing and report generation, improving operational efficiency while reducing human error [15,16]. However, the integration of AI also brings challenges, including concerns about data privacy, potential biases in algorithms, high implementation costs and the risk of job displacement [17]. Innovations such as voice commerce, where purchases are facilitated through voice-activated assistants and Augmented Reality (AR), which provides immersive shopping experiences, are set to transform the industry even further. Predictive commerce, powered by AI systems that anticipate customer needs and blockchain integration for enhanced transaction transparency and trust are also on the horizon.

1.2. Controversies

AI in e-business raises several controversies. These issues are rooted in ethical, social, economic and technical challenges that must be addressed to ensure the responsible use of AI. One of the most prominent concerns is data privacy and security. AI systems in e-business rely heavily on collecting and analyzing customer data to provide personalized experiences and optimize operations. However, this extensive data collection often occurs without explicit consent or sufficient transparency, raising privacy concerns. Moreover, the centralized nature of AI systems makes them vulnerable to data breaches, exposing sensitive customer information. The use of AI to track user behavior and preferences may also be perceived as surveillance, eroding trust between businesses and consumers [18].

Another contentious issue is algorithmic bias and fairness. AI systems may inadvertently perpetuate biases present in their training data, leading to discriminatory practices. For instance, certain customer demographics might be unfairly targeted with specific advertisements or pricing strategies. Compounding this issue is the opaque nature of many AI models, which makes it difficult to identify and rectify biases [19,20]. Additionally, AI-

driven automation in e-business has sparked debates about its impact on employment. By automating routine tasks, AI reduces the need for human intervention. While this increases efficiency, it also raises concerns about job displacement and economic inequality. Furthermore, the ability of AI to predict and influence consumer behavior has led to concerns about manipulative practices. AI-driven personalization may exploit customer vulnerabilities, encouraging impulsive purchases or targeting specific groups with products they may not need [20]. Furthermore, addictive designs, such as gamification and psychological triggers, are increasingly used to foster dependency on e-commerce platforms, raising ethical questions about consumer autonomy.

Accountability is another significant issue. When AI systems make mistakes, such as denying legitimate transactions or producing biased outcomes, it is often unclear who should be held responsible. This ambiguity complicates efforts to address issues of accountability and fairness in e-business. The growing reliance on AI in e-business also introduces systemic risks and reduces human oversight. Overdependence on AI may lead to complacency, where human decision-makers fail to question automated decisions. Furthermore, a malfunction or failure in widely used AI systems may disrupt operations across multiple e-business platforms, causing widespread consequences [21].

In marketing and advertising, AI has sparked debates about ethical boundaries. While personalization enhances customer experiences, hyper-targeted advertising may feel invasive and manipulative [22,23]. Additionally, AI is increasingly used to create deceptive content, such as fake reviews or deepfake advertisements, undermining consumer trust. Moreover, the concentration of AI resources and expertise in a few large corporations creates issues of monopoly and market dominance [24]. Smaller businesses often struggle to compete with the efficiency and personalization capabilities of these larger entities, which may stifle innovation and limit consumer choice.

1.3. Side-Effects Mitigation

Alleviating the side effects of AI in e-business requires a combination of ethical practices, regulatory frameworks, technical innovation and education: (1) Developers should focus on creating diverse and unbiased training datasets to minimize the risk of discriminatory outcomes. Additionally, explainable AI (XAI) tools may improve transparency by providing insights into how AI systems make decisions, allowing for greater accountability and trust [25]. (2) Governments and businesses should collaborate to provide training programs that help workers transition into roles that complement AI-driven automation. Incentives for businesses to retain and retrain their workforce may also encourage a more inclusive transition. (3) Laws should be established to prevent exploitative behaviors, such as using AI to manipulate consumers into making impulsive purchases or to promote addictive designs [26]. (4) Governments need to establish clear rules for accountability when AI systems fail or produce harmful outcomes. Frameworks should outline responsibilities for developers, businesses and operators of AI systems. (5) AI systems should be designed to augment human decision-making rather than replace it entirely. Human-in-the-loop systems ensure that critical decisions remain under human control, reducing the risks associated with overdependence on AI.

1.4. Objective, Research Questions and Contributions

The objective of our research is to investigate the intersection of e-business and AI through a comprehensive analysis of academic publications extracted from the Web of Science Core Collection. It aims to uncover the role of AI in e-business, identify latent themes and highlight emerging research topics.

The main research questions to which the authors intend to answer are as follows:

RQ1: *What are the main topics embedded in the e-business and AI-related publications?*
RQ2: *What are the key topics that could be further studied in the research landscape?*

By employing advanced methodologies such as graphical trend analysis, co-occurrence networks, Sankey diagrams, sentiment analysis and LDA for topic modeling, providing a multidimensional exploration of the field, our research makes significant contributions to both academic and practical domains. It advances knowledge on the role of AI in e-business by providing insights into how AI facilitates business transformation, strategic value creation and operational efficiency. Additionally, the methodological approach, including the use of LDA for topic modeling, enhances bibliometric and text analysis techniques, offering a valuable methodological framework for future research.

The identification of emerging trends and gaps is another key contribution. Practically, this paper underscores the potential of AI in enabling data-driven decision-making, improving information sharing and streamlining business processes. Furthermore, the emphasis on sustainability, particularly through the recent concept of e-agribusiness, demonstrates how businesses can align technological advancements with environmental objectives, promoting sustainable development.

From an academic perspective, this paper enriches the understanding of AI in e-business by identifying four coherent thematic areas using topic modeling. These include (1) data-driven decision-making, (2) ML applications in business innovation, (3) analytical methods for enhancing sustainability, and (4) the role of the semantic web in intelligent systems. This paper also adds to the field of sentiment analysis by comparing methodologies such as custom analysis, VADER and TextBlob, offering insights into their robustness and limitations.

Our research findings may have implications for policymakers and strategists, particularly in guiding digital transformation efforts and fostering sustainable innovations. The dominance of Chinese institutions in the collaborative networks also highlights the geographical and institutional dynamics of this research area, providing a basis for strengthening international collaboration.

Finally, this paper opens pathways for future research by identifying underexplored areas such as the application of AI in e-agribusiness and the semantic web. The future of e-business powered by AI offers transformative potential across several dimensions, promising to revolutionize operations, customer interactions and innovation. It advocates for cross-disciplinary studies that integrate AI, e-business and sustainability. It also identifies new pathways, such as voice and conversational commerce, immersive experiences with AR, emotion-aware AI, blockchain integration, dynamic marketplaces and advanced sentiment analysis.

This research contributes to understanding the role of AI in e-business by conducting a comprehensive bibliometric and textual analysis of academic publications from the Web of Science Core Collection. Our research offers fresh insights into AI's role in e-business. It highlights how AI facilitates business transformation, strategic value creation and operational efficiency in e-business. Key applications include customer personalization, supply chain optimization, fraud detection and marketing automation. Our findings emphasize the integration of e-business principles with environmental objectives, such as the concept of e-agribusiness, demonstrating AI's potential to promote sustainable practices. This research highlights the growing prevalence of positive sentiment in the field and provides a comparative evaluation of sentiment analysis techniques. The dominance of Chinese institutions in the collaboration network provides a basis for strengthening international partnerships in e-business and AI research. The findings offer valuable insights for policymakers and strategists in guiding digital transformation.

Apart from other bibliometric analyses on AI and e-business intersection, our research offers two main contributions:

1. Identification of emerging themes and trends. Using advanced methods like Latent LDA, this research uncovers four coherent thematic areas:
    - Data-driven decision-making in e-business.
    - AI and machine learning applications for innovation and workflow optimization.
    - Analytical methods enhancing sustainability and strategy.
    - The role of the semantic web in intelligent systems and web platforms.
2. Methodological advancements. The integration of LDA topic modeling and sentiment analysis (using custom analysis, VADER and TextBlob) offers a framework for exploring large text corpora, providing actionable insights for future bibliometric studies.

The remainder of this paper is organized as follows: Section 2 describes the most recent and relevant literature review and Section 3 presents the methodology, including the data sources, analytical tools and techniques used. Section 4 presents the results, detailing the key findings. Section 5 provides a discussion on future pathways, exploring their significance and potential applications. Section 6 concludes this paper by summarizing the main contributions, highlighting limitations and suggesting directions for future research.

2. Literature Review

AI has become a widely adopted technology in the e-commerce sector, enabling businesses to enhance various aspects of their operations. Despite the growing use of AI in this industry, there has been limited bibliometric analysis conducted to systematically map the research in this field. While AI applications in e-commerce continue to expand, a comprehensive overview of scholarly work that tracks the progress, key research areas and emerging trends has not been extensively explored, leaving a gap in understanding the full scope of academic contributions to AI in e-commerce.

A bibliometric analysis and a literature review were combined to explore AI research in e-commerce, analyzing 4335 documents and 229 key articles [27]. It highlighted recommender systems, sentiment analysis, trust, personalization and optimization as central themes, with China leading in research contributions. Another piece of research examined the impact of AI on fashion e-commerce through a systematic review of 219 publications from Web of Science and Scopus [28]. The articles were categorized using AI techniques into two groups, enabling the identification of research gaps in AI applications. These gaps highlighted opportunities for future exploration. Furthermore, 711 articles on AI and Big Data Analytics (BDA) published between 2012 and 2022 were reviewed [29], identifying key trends and contributions across multiple disciplines. Data from 59 countries showed the USA as the most influential nation (101 documents; 5405 citations) and China as the most productive (204 documents; 2371 citations). Symbiosis International University in India emerged as the leading institution. This research highlighted five main clusters: Business and Management, Engineering, Healthcare, Sustainable Operations and Industry 4.0 and Tourism and Hospitality, with applications in real-world challenges. Prominent research areas were highlighted. A total of 658 publications from Scopus on digitalization between 2018 and 2022 were further analyzed using RStudio and VOSviewer [30]. It identified ten key application areas, including Industry 4.0, AI in supply chain management, augmented reality, healthcare and digital collections.

Another piece of research reviewed research on how technologies such as NLP, sentiment analysis, business intelligence and IoT data influenced customer behavior in online and virtual marketplaces [31]. A quantitative review of 84 studies from 2022 narrowed down to 17 empirical sources using tools like PRISMA, AXIS and VOSviewer. The find-

ings showed that predictive algorithms, data visualization and artificial neural networks enhanced performance in livestreaming e-commerce by shaping purchasing habits, loyalty and customer engagement. It highlighted the significant role of AI and data-driven tools in transforming the retail metaverse. Moreover, [32] conducted a bibliometric analysis of AI in e-commerce from 2014 to mid-2023 using data from 669 manuscripts in the Scopus database. It evaluated author status, nations, affiliations and keywords using VOSviewer and R Studio. The UK emerged as the leading publisher with the strongest co-authorship links, while the most productive journal was Proceedings of the International Conference on Electronic Business. Key research areas included business, management and accounting, with popular keywords such as ML, e-commerce, recommender systems, fraud detection and data mining. AI has driven rapid changes, enhancing efficiency, productivity and cost-effectiveness for companies. Additionally, [33] analyzed research trends in AI applications within e-commerce through systematic review, network analysis and bibliometric methods, focusing on co-authorship, co-citation and keyword analysis. The findings revealed a growing academic interest in AI's role in e-commerce, highlighting its importance for theoretical and applied research. Another piece of research analyzed trends and developments in e-retailing, focusing on Indonesia and the global context, using bibliometric analysis and systematic literature review with VOSviewer software [34]. The findings identified eight key clusters in e-retailing: chatbots, customer satisfaction, purchase intention, customer experience, artificial intelligence, social media and mobile commerce.

The impact of AI in marketing through a bibliometric and intellectual network analysis of 1580 papers published between 1982 and 2020 was further reviewed [35]. Using methods such as co-citation, co-occurrence analysis and the Louvain algorithm, the researchers identified key authors, sources, sub-themes and future directions for AI in marketing. Practical examples highlighted AI's transformative role in e-commerce, particularly through Amazon's use of AI to enhance customer experiences and streamline online shopping. Another piece of research examined the influence of AI on product recommendations in e-commerce, focusing on its impact on consumer behavior and purchasing decisions [36]. Using bibliometric analysis of 103 articles from 2017 to 2023 in the Scopus database, 29 relevant studies were analyzed. The findings revealed that AI significantly affects recommendation systems, decision support systems, customer behavior and customer trust, highlighting the role of product recommendations in shaping e-commerce outcomes. However, the review identified a limited number of studies addressing the implementation of AI in areas such as consumer behavior, trust and purchasing decisions, underscoring the need for further research in this field. The challenges of integrating AI into businesses were investigated [37], noting that 85% of AI initiatives fail to meet expectations. Using Alibaba's e-commerce fulfillment center as a case study, it analyzed how AI technology, human resources and processes can be managed to create value. Key AI resources, including data, algorithms and robots, were found to generate capabilities like forecasting, planning and learning when effectively orchestrated with facilities and information systems.

Another piece of research examined the applications of ML and AI in e-commerce, corporate management and finance [38]. AI has been employed to enhance customer experience, optimize supply chain management, improve operational efficiency and maintain low costs while ensuring product quality. Key applications included sales growth, profit maximization, sales forecasting, inventory management, security, fraud detection and portfolio management. ML and deep learning were highlighted as significant tools for analyzing complex and diverse data, with ongoing developments targeting industries such as food, finance and e-commerce to drive innovation and efficiency. How AI can optimize digital marketing strategies through a systematic literature review using the PRISMA framework was also analyzed [39]. Investigating 211 articles, this research identified key clusters:

AI/ML algorithms, social media, consumer behavior, e-commerce, digital advertising, budget optimization and competitive strategies. The findings highlighted AI's transformative potential in enhancing marketing efforts and provided insights into trends, challenges and opportunities at the intersection of AI and digital marketing. Reference [40] examined the impact of different types of online customer service: AI, manual and human–machine collaboration, on customer purchase intentions, using service encounter and superposition theories. Two shopping experiments revealed that perceived service quality significantly influences purchase intention, serving as a mediating factor, while product type moderates this relationship. Notably, human–machine collaboration demonstrated a superposition effect, enhancing customer purchase intention.

Technologies like ML and deep learning have enhanced e-commerce operations through personalized product recommendations, automated customer service, optimized search interfaces and improved security [41]. These AI-driven solutions have enabled companies to enhance customer experiences, lower operational costs and increase efficiency, supporting the continued evolution of e-commerce. Another piece of research explored the implementation of AI fairness in e-commerce, focusing on coupon allocation based on clickstream sessions [42]. While AI fairness successfully reduced disparate outcomes and met fairness requirements with minimal impact on prediction performance, it also increased financial costs. The findings highlighted the trade-offs between ethical AI practices and operational expenses, offering insights into designing information systems that prioritize fairness while balancing performance and cost considerations.

The role of AI in e-commerce was examined [43], emphasizing its transformative potential for personalization and product customization. However, it highlighted the ethical challenges surrounding the explainability of AI systems. This research explored the concept of XAI, a field focused on making ML models transparent and interpretable. The findings underscored the need to improve AI models to enhance their comprehensibility, laying a foundation for developing XAI frameworks that address ethical concerns while maintaining effectiveness in e-commerce applications. Nonetheless, the future of AI in e-commerce is marked by greater personalization, automation and convenience for customers, alongside enhanced efficiency and profitability for businesses. The researchers found that AI has already transformed business operations and customer experiences, enabling companies to reduce costs and improve performance [44]. Table 1 provides a comparative view of the previous studies.

Table 1. Comparation of the previous research on e-business and AI.

Ref	Objective	Methods	Keywords	Main Findings
[27]	Synthesizes research on AI in e-commerce and proposes IS research guidelines.	Bibliometric analysis and extensive literature review (4335 documents analyzed, 229 articles reviewed).	AI, E-commerce, Recommender Systems, Personalization, Sentiment Analysis, Trust, Optimization.	Recommender systems dominate research. Identified core themes: sentiment analysis, trust, personalization, optimization. China is leading in research output.
[28]	Examines AI's significance and impact on fashion e-commerce.	Systematic review using Web of Science and Scopus databases (219 publications).	Fashion E-commerce, AI, Research Gaps.	Identified two categories in AI's application in fashion e-commerce and research gaps for future exploration.

Table 1. *Cont.*

Ref	Objective	Methods	Keywords	Main Findings
[29]	Summarizes bibliometric articles on AI and BDA across multiple fields.	Bibliometric review (711 documents from Scopus Q1 and Q2 journals, 2012–2022).	AI, BDA, Bibliometric Analysis, NLP, Deep Learning.	USA is most influential, China is most productive. Five main disciplines for AI/BDA: Business, Engineering, Healthcare, Sustainable Operations, Tourism.
[30]	Analyzes publications on digitalization from 2018 to 2022.	Bibliometric review using RStudio and VOSviewer (658 documents).	Digitalization, Big Data, AI, Industry 4.0.	Identified ten fields of digitalization applications, with AI and supply chain management being prominent areas.
[31]	Examines AI tools' role in retail metaverse.	Quantitative literature review with PRISMA framework (17 articles analyzed).	Retail Metaverse, AI, Sentiment Analysis, Predictive Algorithms.	AI enhances business performance in live-streaming e-commerce via predictive algorithms, sentiment analysis, and data visualization tools.
[32]	Assesses AI research landscape and trends in e-commerce (2014–2023).	Bibliometric analysis using Scopus database, VOSviewer, and R Studio (669 manuscripts).	E-commerce, AI, ML, Fraud Detection.	Popular areas: business, management, and accounting. UK leads in publications and co-authorship. Key terms: recommender systems, fraud detection, decision-making systems.
[33]	Analyzes research trends on AI in e-commerce.	Systematic review, network, and bibliometric analysis.	AI, E-commerce, Productivity, Efficiency.	AI use in e-commerce is increasing, focusing on efficiency, cost reduction, and productivity improvement.
[34]	Analyzes trends and developments in e-retailing, focusing on Indonesia and globally.	Bibliometric analysis and literature review using VOSviewer.	E-retailing, Chatbots, Customer Satisfaction, Mobile Commerce.	Chatbots, customer satisfaction, and purchase intention are key to e-retailing success.
[35]	Comprehensive review of AI in marketing (1982–2020).	Bibliometric, abstract, and intellectual network analysis (1580 papers).	AI, Marketing, E-commerce, Amazon.	Identified research sub-themes and practical examples of AI in e-commerce, with Amazon as a key example.
[36]	Explores AI's impact on product recommendations in e-commerce.	Bibliometric analysis of 103 articles from Scopus (2017–2023).	AI, Product Recommendations, Customer Trust, Decision-making.	AI influences consumer behavior, customer trust, and purchasing decisions in e-commerce.
[37]	Analyzes AI's success factors in e-commerce using Alibaba as a case study.	Resource orchestration perspective.	AI, Resource Management, E-commerce, Alibaba.	AI resources (data, algorithms, robots) co-evolve with human capabilities to drive efficiency and effectiveness.
[38]	Discusses AI applications in e-commerce, corporate management, and finance.	Literature review.	AI, ML, Deep Learning, Fraud Detection.	AI improves customer experience, operational efficiency, and business profitability.
[39]	Optimizes digital marketing strategies using AI.	Systematic literature review with PRISMA framework (211 articles).	AI, Digital Marketing, Consumer Behavior, Social Media.	AI/ML algorithms are crucial for digital marketing, focusing on personalization, social media, and competitive strategies.
[40]	Examines AI's role in online customer service and purchase intention.	Shopping experiments based on service encounter and superposition theories.	AI, Customer Service, Purchase Intention.	Human–machine collaboration enhances perceived service quality and purchase intention.
[41]	Examines AI's transformative role in e-commerce.	Descriptive analysis.	AI, Personalization, Automation, E-commerce.	AI enables personalization, reduces costs, and enhances efficiency in e-commerce.

Table 1. *Cont.*

Ref	Objective	Methods	Keywords	Main Findings
[42]	Explores AI fairness in e-commerce.	Developed AI fairness model using clickstream sessions.	AI Fairness, E-commerce, Clickstream Sessions.	AI fairness increases financial costs but ensures compliance with fairness requirements.
[43]	Investigates XAI models' role in e-commerce.	Word cloud, voyance, and concordance analysis.	Explainable AI, XAI, E-commerce, Personalization.	Proposed XAI models for transparent and interpretable AI systems in e-commerce.
[44]	Analyzes AI's evolving role in e-commerce.	Qualitative analysis.	AI, E-commerce, Personalization, Automation.	AI enhances personalization, automation, and profitability in e-commerce.

The key gaps identified in the literature are the lack of analysis on latent topics within academic publications and the fresher exploration of emerging trends and opportunities that AI could bring to e-business. While existing studies have focused on specific applications or case studies, there remains a need for a comprehensive approach to uncover hidden themes, research trajectories and potential future directions in the integration of AI within e-business ecosystems.

3. Methodology

This section outlines the input dataset and the proposed methodology, detailing the steps taken to derive insights from the data. The selected methods, Latent Dirichlet Allocation, sentiment analysis, visual representations, graphical networks, trends, distributions and frequency analysis, are employed to achieve the research objectives and address the research questions.

3.1. Input Dataset

The dataset (https://github.com/simonavoprea/ebai) is downloaded from Web of Science (WoS) Core Collection previously filtered by "e-business" and "artificial intelligence", aiming to analyze the publications in which these two topics are studied together. 448 of full records were found by mid-November 2024 (17th of November). The search focused on the two concepts, "e-business" and "artificial intelligence", aiming to explore their intersection and derive insights from academic publications. All publication types, such as C (Conference papers or proceedings), J (Journal articles) and S (Short papers, symposium contributions, or other specific formats), were extracted. No additional filters were applied to the search criteria. No full text was obtained, only the abstract was available for each publication. Out of the entire set of columns (over 70), we keep the most relevant ones (as in Table 2):

Table 2. Relevant columns in the input dataset, their meanings, along with examples of possible values.

Column Name	Meaning	Example Value
Publication Type	Type of publication (e.g., journal, conference)	"C" (Conference), "J" (Journal)
Article Title	Title of the research article	"The Impact of Artificial Intelligence and Internet of Things on E-business"
Source Title	Name of the journal or conference where the article was published	"ANNALS OF OPERATIONS RESEARCH"
Document Type	Specific type of the document	"Proceedings Paper", "Article"

Table 2. *Cont.*

Column Name	Meaning	Example Value
Conference Date	Date of the conference	"26 April 2019", "NaN"
Conference Location	Location of the conference	"Charlottesville, VA", "NaN"
Author Keywords	Keywords provided by the authors	"Artificial Intelligence; E-business; IoT"
Keywords Plus	Additional keywords generated by indexing	"E-COMMERCE; AI", "PREDICTION"
Abstract	Summary of the article	"This study explores the impact of Artificial Intelligence..."
Affiliations	Affiliations of the authors	"Tshwane University of Technology", "NaN"
Times Cited, All Databases	Total citations from all databases	-
180 Day Usage Count	Number of accesses in the last 180 days	-
Since 2013 Usage Count	Number of accesses since 2013	-
Publisher	Publisher of the journal or conference proceedings	"IEEE", "SPRINGER"
Publisher City	City where the publisher is based	"NEW YORK", "BERLIN"
Publisher Address	Address of the publisher	"345 E 47TH ST, NEW YORK, NY 10017 USA"
Publication Date	Specific date of publication	"JUL", "NaN"
Publication Year	Year of publication	2019, 2023
WoS Categories	Web of Science categories for the article	"Computer Science, Interdisciplinary Applications"
Research Areas	Research areas associated with the article	"Computer Science; Engineering"

Summary statistics for the non-numerical columns and numerical columns are provided in Tables 3 and 4.

Table 3. Summary statistics for the non-numerical columns.

Column	Count	Unique	Top Value	Frequency
Publication Type	448	3	C	282
Article Title	448	448	The Impact of Artificial Intelligence and Internet of Things in the Transformation of E-Business Sector	1
Source Title	448	293	UNCERTAINTY REASONING FOR THE SEMANTIC WEB I	22
Document Type	448	7	Proceedings Paper	282
Conference Date	331	190	11–15 November 2007	27
Conference Location	331	149	Busan, SOUTH KOREA	27
Author Keywords	335	334	mash up; drag and drop; faceted browsing; simultaneous editing; ontology alignment; end-user programming	2
Keywords Plus	188	178	ALGORITHM	4
Abstract	445	445	[Full abstract text]	1
Affiliations	408	341	University of Salento	5
Publisher	448	73	SPRINGER-VERLAG BERLIN	120

Table 3. *Cont.*

Column	Count	Unique	Top Value	Frequency
Publisher City	448	59	BERLIN	120
Publisher Address	448	81	HEIDELBERGER PLATZ 3, D-14197 BERLIN, GERMANY	120
Publication Date	105	50	AUG	9
WoS Categories	447	138	Computer Science, Artificial Intelligence	82
Research Areas	447	65	Computer Science	234

Table 4. Summary statistics for the numerical columns.

Column	Count	Mean	Std Dev	Min	25%	50%	75%	Max
Cited Reference Count	448	30.801	31.733	0	12	20	36	244
Times Cited, WoS Core	448	7.772	19.220	0	0	1	6	197
Times Cited, All DBs	448	8.391	21.211	0	0	1.5	7	231
180 Day Usage Count	448	3.002	10.404	0	0	0	1	135
Since 2013 Usage Count	448	17.781	42.799	0	1	3	12.25	442
Publication Year	448	2012	7.466	1999	2006	2010	2020	2025

3.2. Methodology Steps

The proposed methodology, depicted in the flowchart in Figure 1, outlines a systematic approach for analyzing research trends and sentiment related to "e-business" and "artificial intelligence" by leveraging bibliometric and text analysis techniques. A detailed breakdown of each step in the methodology is offered in the following paragraphs.

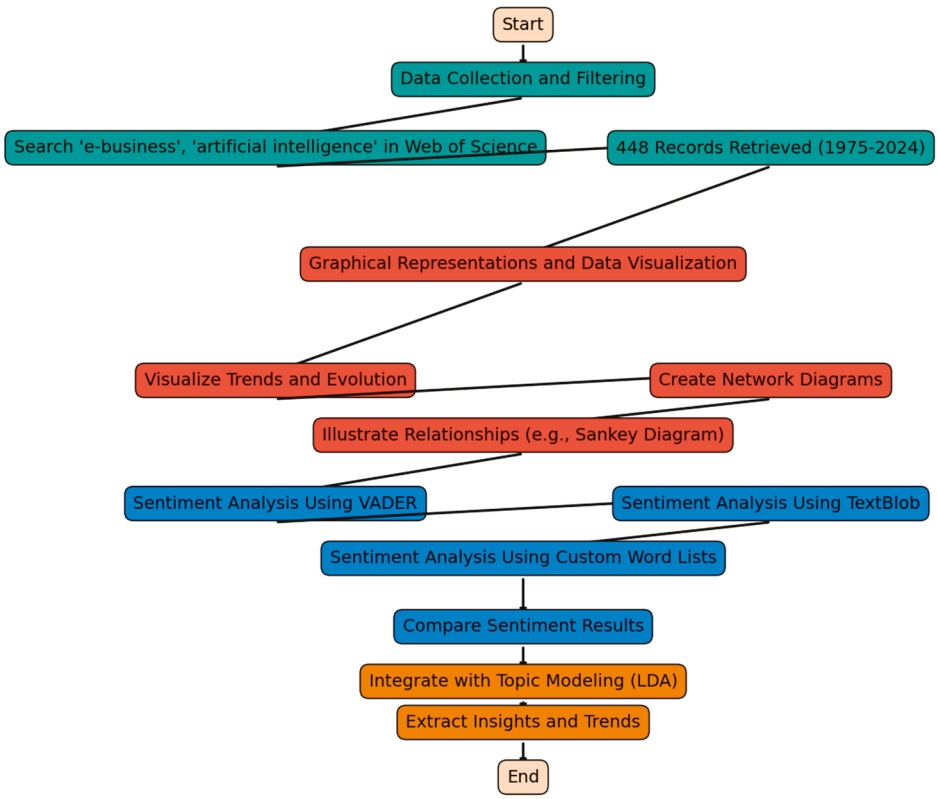


Figure 1. Flowchart of the proposed methodology.

Step 1. Data collection and filtering.

This research begins with the identification and retrieval of relevant literature from the Web of Science database. Keywords such as "e-business" and "artificial intelligence" are used to filter records, spanning from 1999 to 2024, yielding a dataset of 448 records.

Step 2. Graphical representations and data visualization.

The collected data undergo processing for visualization to make trends and patterns more understandable. Two main visualization objectives are pursued:

- Visualizing trends and evolution. This involves identifying temporal changes in research activity, such as the frequency of publications over time, key contributing institutions or thematic shifts.
- Creating network diagrams. Network diagrams depict relationships between keywords or research topics, highlighting interconnected areas of study.

Step 3. Illustrate relationships (e.g., Sankey diagram).

Relationships among key terms, concepts or research trends are illustrated using Sankey diagrams. These diagrams show the flow or connections between key elements, such as those focused on "business", publication years and research areas, offering a clear picture of how different ideas or areas interact or temporarily evolve.

Step 4. Sentiment analysis.

Sentiment analysis is performed to assess the emotional tone and context of the literature:

- Using VADER. A rule-based sentiment analysis tool tailored for text, focusing on the positive, negative or neutral tone of the text (opinions).
- Using TextBlob. A tool for NLP that provides a polarity score (positive/negative) and subjectivity score.
- Using custom word lists. A tailored approach where domain-specific terms are manually curated, assigning sentiment scores based on the context of e-business and artificial intelligence.

Step 5. Compare sentiment results.

Results from all sentiment analysis methods are compared to identify patterns or discrepancies. This triangulation ensures consistent findings by capturing both general and context-specific sentiments.

Step 6. Integrate with topic modeling.

Sentiment analysis results are integrated with LDA topic modeling to identify dominant research themes and their sentiment context. This step links the emotional tone of discussions to specific themes or areas within the dataset, offering deeper insights. The final step synthesizes the findings, highlighting key trends, themes and patterns in the research landscape of e-business and artificial intelligence.

The combination of sentiment analysis, topic modeling and visualizations creates a holistic understanding of the field. Using multiple sentiment tools (VADER, TextBlob, custom word lists) ensures both breadth and depth, accommodating general sentiment and domain-specific nuances. Visualizing relationships through Sankey diagrams and network diagrams enhances the interpretability of complex connections between concepts.

For the application of the LDA to abstracts of publications in the WoS dataset, each abstract consisting of 200–350 words, we define the probabilistic generative process and associated mathematical equations.

Step 6.1. Key notations are defined in Table 5.

Table 5. LDA model variables.

Variable	Depiction
D	Number of abstracts (documents) in the dataset
V	Vocabulary size (total unique words across all abstracts)
K	Number of topics k.
$\boldsymbol{N_d}$	Number of words in document d (approximately 200–350 words in this context)
$\boldsymbol{w_{d,n}}$	The n-th word in document d, where $w_{d,n} \in \{1, 2, \ldots, V\}$
$\boldsymbol{\theta_d}$	Topic distribution for document d, a K-dimensional vector
$\boldsymbol{\phi_k}$	Word distribution for topic k, a V-dimensional vector
$\boldsymbol{z_{d,n}}$	Latent topic assignment for word $w_{d,n}$, where $z_{d,n} \in \{1, 2, \ldots, K\}$

Step 6.2. Generative process

LDA models the generative process for each word in the abstracts as follows:

1. For each document d in the corpus—sample a topic distribution θ_d from a Dirichlet distribution:

$$\theta_d \sim Dirichlet(\propto) \tag{1}$$

where α is a K-dimensional hyperparameter controlling the sparsity of θ_d.

2. For each topic k—sample a word distribution ϕ_k from another Dirichlet distribution:

$$\phi_k \sim Dirichlet(\beta) \tag{2}$$

where β is a V-dimensional hyperparameter controlling the sparsity of ϕ_k.

3. For each word $w_{d,n}$ in document d, perform the following:
   - Sample a topic assignment $z_{d,n}$ from the document's topic distribution:

$$z_{d,n} \sim Categorial(\theta_d) \tag{3}$$

   - Sample the word $w_{d,n}$ from the topic's word distribution:

$$w_{d,n} \sim Categorial\left(\phi_{z_{d,n}}\right) \tag{4}$$

Step 6.3. Joint distribution. The joint probability of the observed words w, topic assignments z and latent variables θ and ϕ is given by the following:

$$P(w, z, \theta, \phi|\alpha, \beta) = \prod_{k=1}^{K} P(\phi_k|\beta) \prod_{d=1}^{D} P(\theta_d|\alpha) \prod_{n=1}^{N_d} P(z_{d,n}|\theta_d) P\left(w_{d,n}\middle|\phi_{z_{d,n}}\right) \tag{5}$$

Step 6.4. Marginal likelihood. The likelihood of the observed data (words w) is obtained by marginalizing over z, θ and ϕ:

$$P(w|\alpha, \beta) = \iint \sum_{z} P(w, z, \theta, \phi|\alpha, \beta) d\phi d\theta \tag{6}$$

Step 6.5. Approximation methods

- Variational inference. Approximate $P(w, z, \theta, \phi)$ with a variational distribution $q(\theta, \phi, z)$ to minimize the Kullback–Leibler (KL) divergence:

$$KL(q||\ P) = \mathbb{E}_q[\log q(\theta, \phi, z) - \log P(\theta, \phi,\ z|w, \alpha, \beta)] \tag{7}$$

- Gibbs sampling—use collapsed Gibbs sampling to iteratively sample the topic assignments $z_{d,n}$ based on the conditional distribution:

$$P(z_{d,n} = k|w,\ z_{-d,-n}, \propto, \beta)\alpha \frac{n_{d,k}^{-n} + \alpha}{n_{d,.}^{-n} + K\alpha} \frac{n_{k,w}^{-d,-n} + \beta}{n_{k,.}^{-d,-n} + V\beta} \tag{8}$$

where $n_{d,k}$—count of words in document d assigned to topic k.

$n_{k,w}$—count of word w assigned to topic k.

The superscript $-n$ or $-d, -n$ indicates the exclusion of the current word or document.

Step 6.6. Evaluation metrics

- Perplexity (measures the generalization ability of the model):

$$Perplexity = exp\left(-\frac{\sum_{d=1}^{D} \log P(w_d|\alpha, \beta)}{\sum_{d=1}^{D} N_d}\right) \tag{9}$$

- Topic coherence (measures interpretability of topics):

$$Coherence = \sum_{i<j} \log \frac{P(w_i, w_j)}{P(w_i)P(w_j)} \tag{10}$$

where $P(w_i, w_j)$ is the probability of words w_i and w_j co-occurring in the same document.

For abstracts in the WoS dataset, the LDA reveals thematic structures by (1) treating each abstract as a document and processing the text into a bag-of-words representation; using variational inference or Gibbs sampling to estimate topic distributions (θ_d) for each abstract; (2) extracting insights such as the dominant research topics, their evolution over time and key trends by analyzing the topic-word distributions (ϕ_k) and document-topic distributions (θ_d).

4. Results

4.1. Graphical Representations

Figure 2 illustrates the number of publications related to e-business and AI over time, showing distinct trends across different periods. Between 2000 and 2010, there was a steady rise in publications, peaking around 2005–2010, reflecting the initial integration of AI technologies into e-business and the growing research interest in their applications. Following this peak, the number of publications declined slightly and entered a relatively stable phase during the period from 2011 to 2018. This phase indicates a period with fewer groundbreaking developments. A renewed interest is evident post-2018, with a sharp increase in publications around 2020 and beyond, reaching a new peak. This resurgence can be attributed to advancements in AI technologies, including ML, deep learning, blockchain and big data analytics, as well as the rapid expansion of e-commerce during the COVID-19 pandemic.

The violin plot in Figure 2 represents the distribution of citation counts for publications across different years. Publications from earlier years, particularly pre-2005, show a wider spread in citation counts, with some papers receiving significantly high citations (above 200). During the mid-years, between 2005 and 2015, the citation distribution narrows, suggesting fewer outliers with extremely high citations. Most publications during this period received moderate and consistent citation counts. In the post-2015 period, recent publications exhibit lower citation counts with tighter distributions.

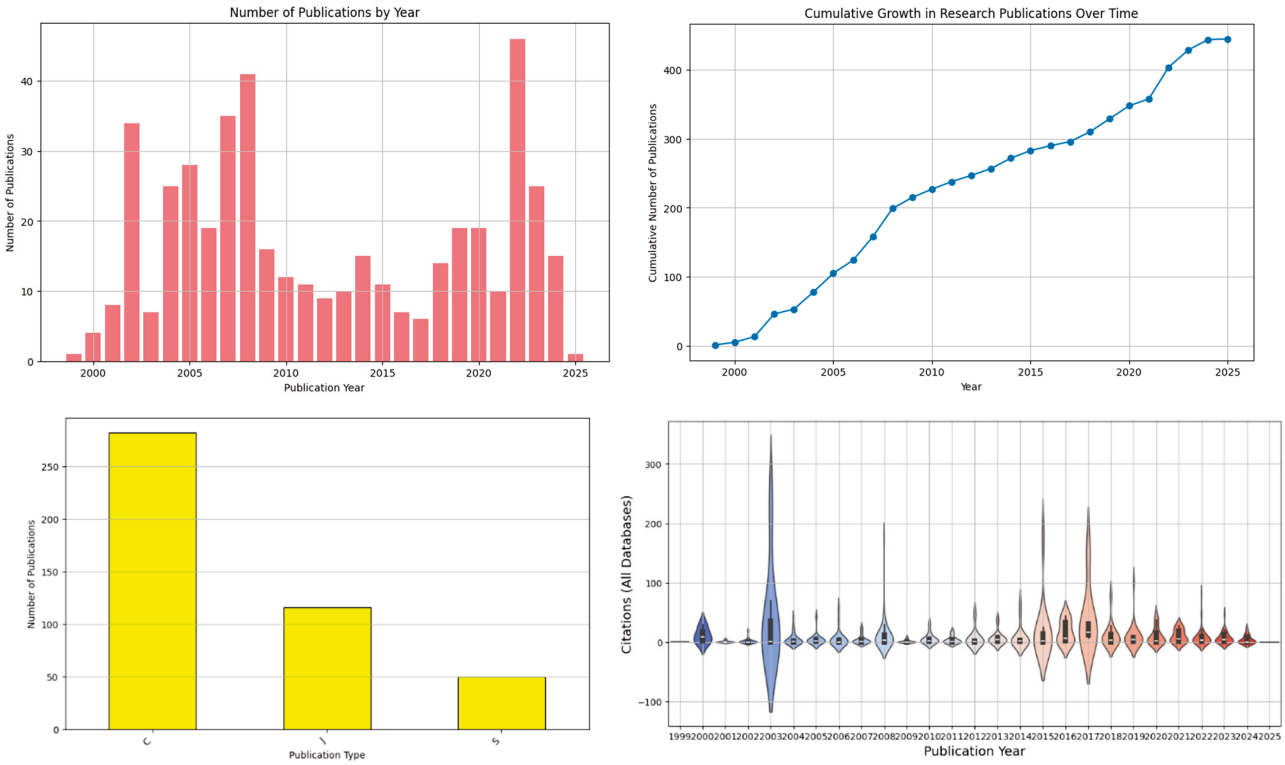


Figure 2. Number and cumulative growth in research publications, type and distribution of citations.

The two plots in Figure 3 provide a detailed overview of the research focus on "e-business" and "artificial intelligence" by examining keyword frequency and the distribution of publications across research areas. The first plot highlights the most commonly used keywords. Significant terms, such as "e-commerce", "machine learning" and "semantic web" suggest a strong emphasis on integrating AI technologies into digital business processes and structuring data for meaningful applications. Additionally, terms like "trust" and "digitalization" indicate emerging themes, reflecting growing attention to the ethical, relational and transformational aspects of AI in business contexts. Keywords like "data mining" and "classification" reveal the technical approaches frequently employed in this research.

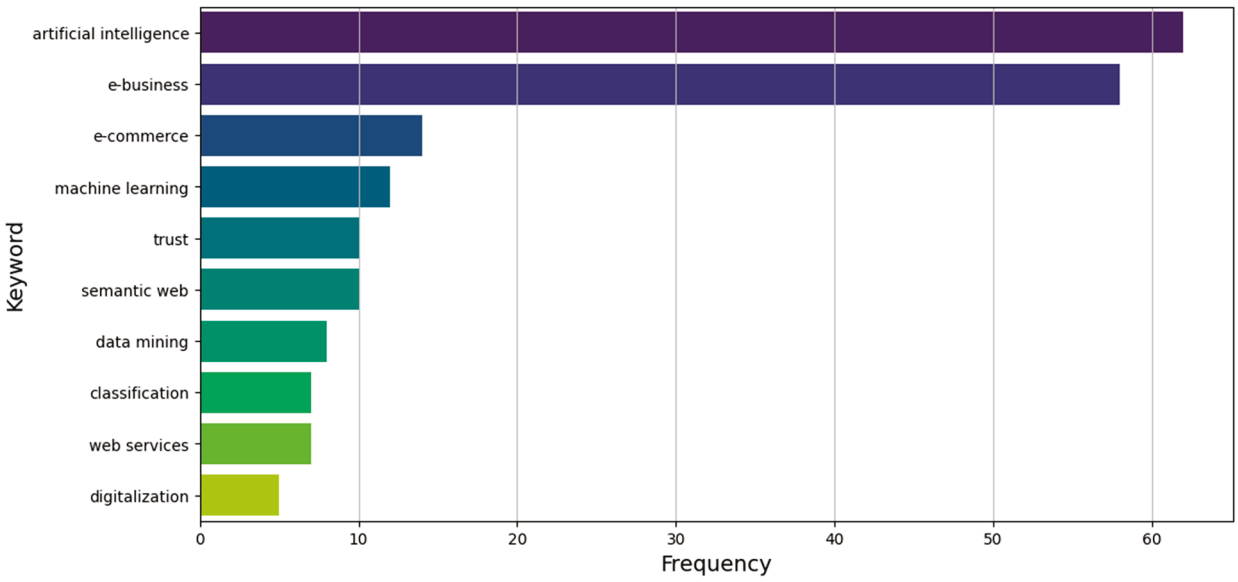


Figure 3. *Cont.*

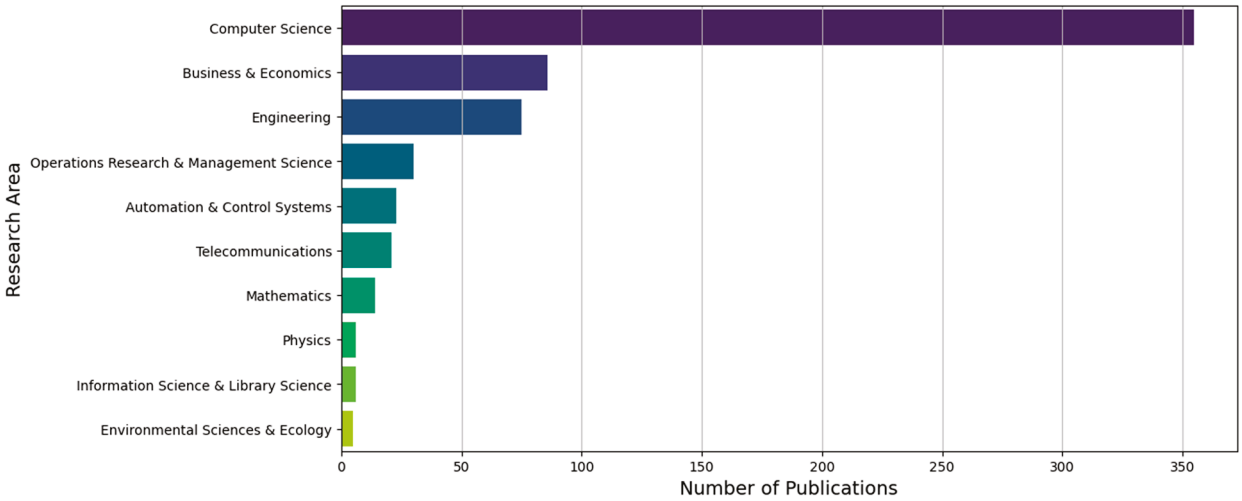


Figure 3. Top ten most common keywords and the number of publications by research area (top ten).

The second plot focuses on the distribution of publications across research areas, showing that the majority of work is concentrated in "Computer Science". "Business & Economics" is another major focus, highlighting the application of AI in business strategies, decision-making and economic modeling. Significant contributions also come from "Engineering" and "Operations Research & Management Science". In Figure 4, the word cloud reflects the expected output.

Figure 4. Word cloud on the Author Keywords column.

Figure 5 illustrates the collaboration network among the top institutions engaged in e-business and AI. The Nanjing University of Finance & Economics is at the center of the network, acting as a hub with connections to multiple other institutions, including the Chinese Academy of Sciences, Nanjing University and the University of Technology Sydney.

Figure 6 shows the relationships between key terms in prosumer research based on their co-occurrence in the literature. "E-business" is the central node, connected to numerous keywords like "artificial intelligence", "e-commerce", "trust" and "machine learning". Strong connections exist between "artificial intelligence" and "e-commerce", as well as between "machine learning" and "artificial intelligence". Keywords like "ontology", "web services" and "classification" have relatively fewer connections.

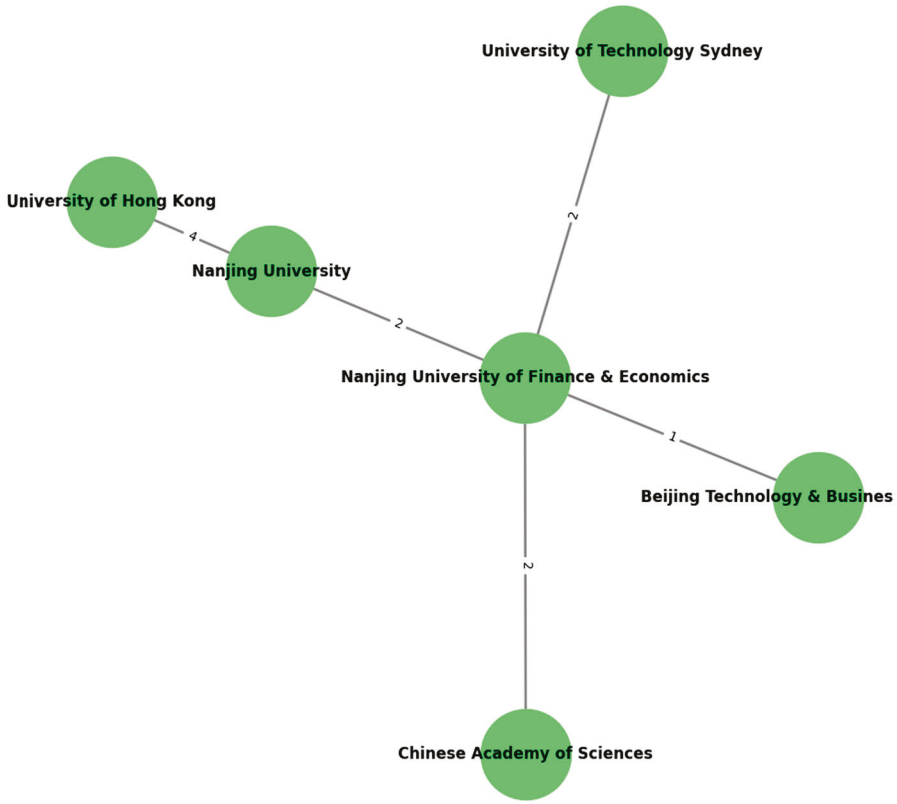


Figure 5. Collaboration network of top seven institutions in e-business and AI. The numbers represent collaborations between universities and institutes.

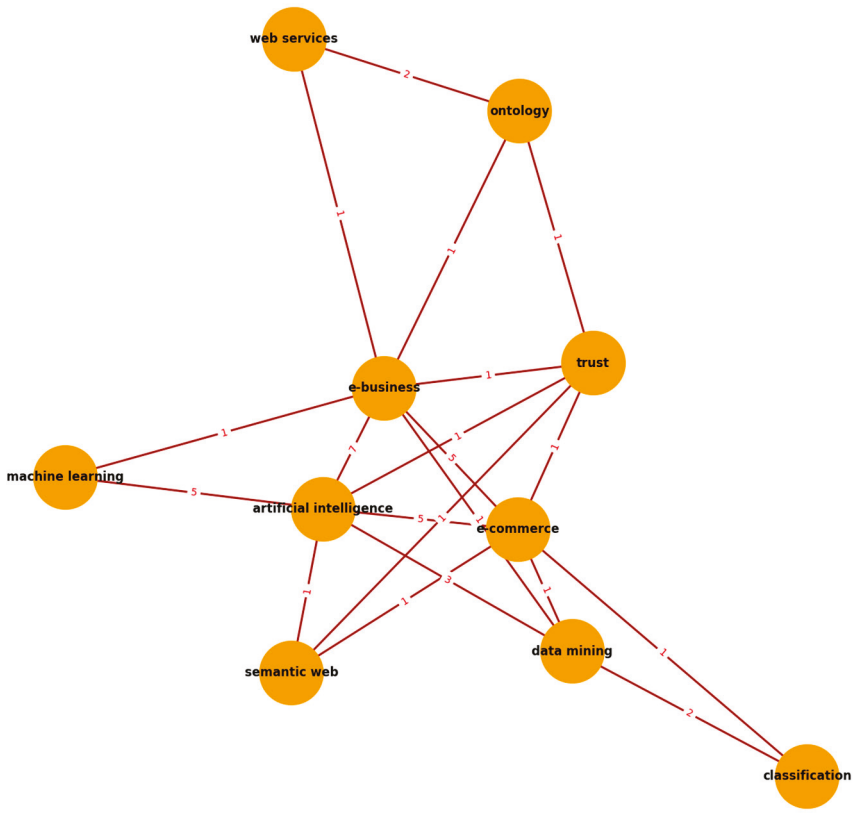


Figure 6. Co-occurrence of keywords in e-business and AI publications. The numbers are the co-occurrences where two keywords were found.

At the center of the network in Figure 7, "artificial intelligence", "e-business" and "business & economics" act as critical connecting points, linking various research areas like "computer science", "engineering" and "operations research & management science". Keywords such as "semantic web", "trust" and "machine learning" are closely connected to both technical disciplines and business-focused research areas, reflecting their importance in technical innovation and business applications. Emerging fields like "environmental sciences & ecology" and "information science & library science" are less connected, suggesting that their integration with e-business and AI is still developing.

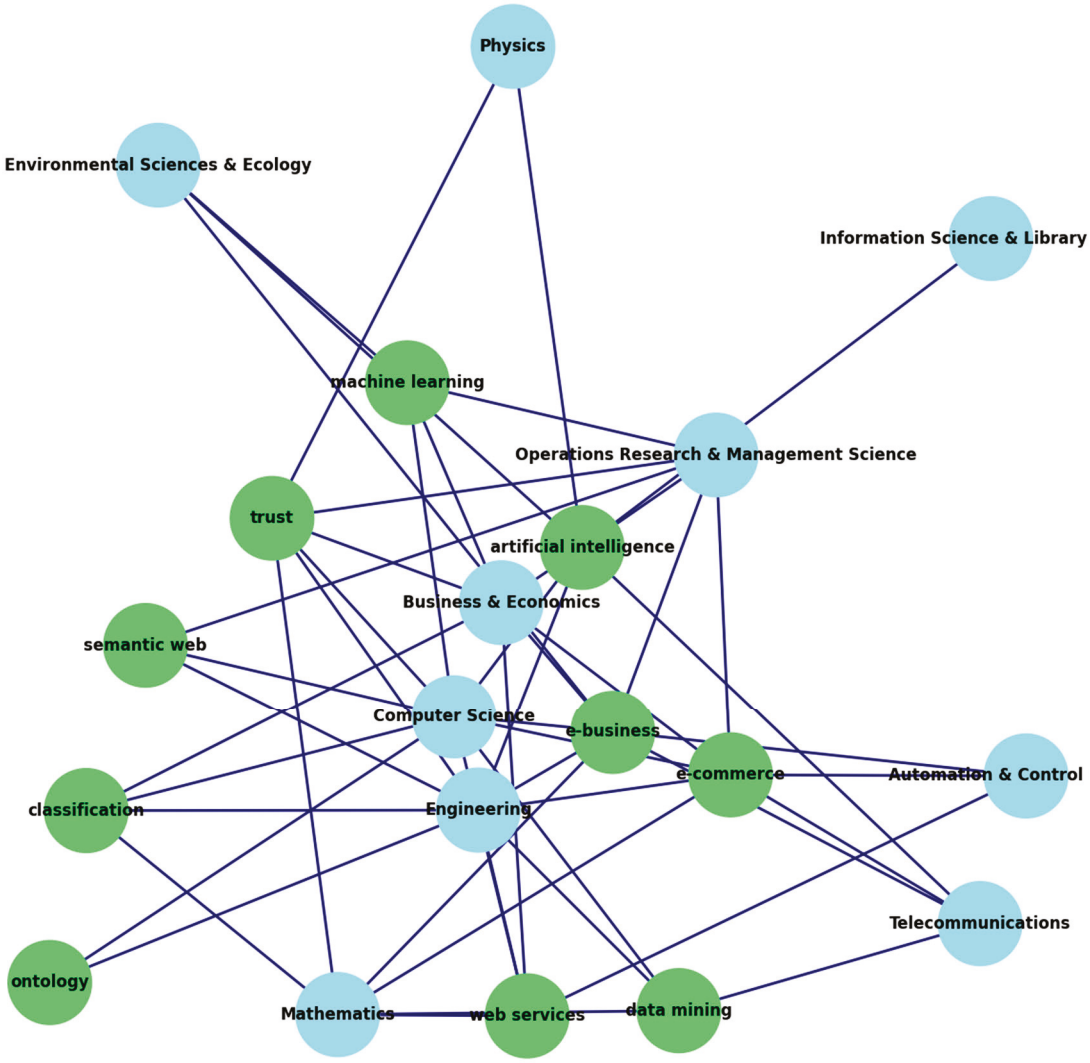


Figure 7. Network of research areas and keywords in e-business and AI.

The Sankey diagram in Figure 8 illustrates the connections between keywords, research areas and publication years for business-related research. Keywords such as "business transformation", "business value" and "e-business strategy" are prominent, contributing significantly to areas like "Operations Research & Management Science". Additionally, the keyword "e-agribusiness" appears connected to "Environmental Sciences & Ecology", indicating the application of e-business principles in sustainable agricultural practices. The temporal distribution, spanning from 2018 to 2024, shows a notable diversification in research over time. Earlier studies, particularly from 2018 to 2020, were predominantly tied to established fields like "Business & Economics" and "Computer Science." In contrast, research from 2021 onward reveals increasing interdisciplinary engagement, with growing connections to fields such as "Environmental Sciences & Ecology" and "Mate-

rials Science." This shift highlights a broadening focus on sustainability and innovative applications. The emergence of topics like "business transformation" and "business analytics" in recent years reflects a greater emphasis on digital transformation and data-driven decision-making processes.

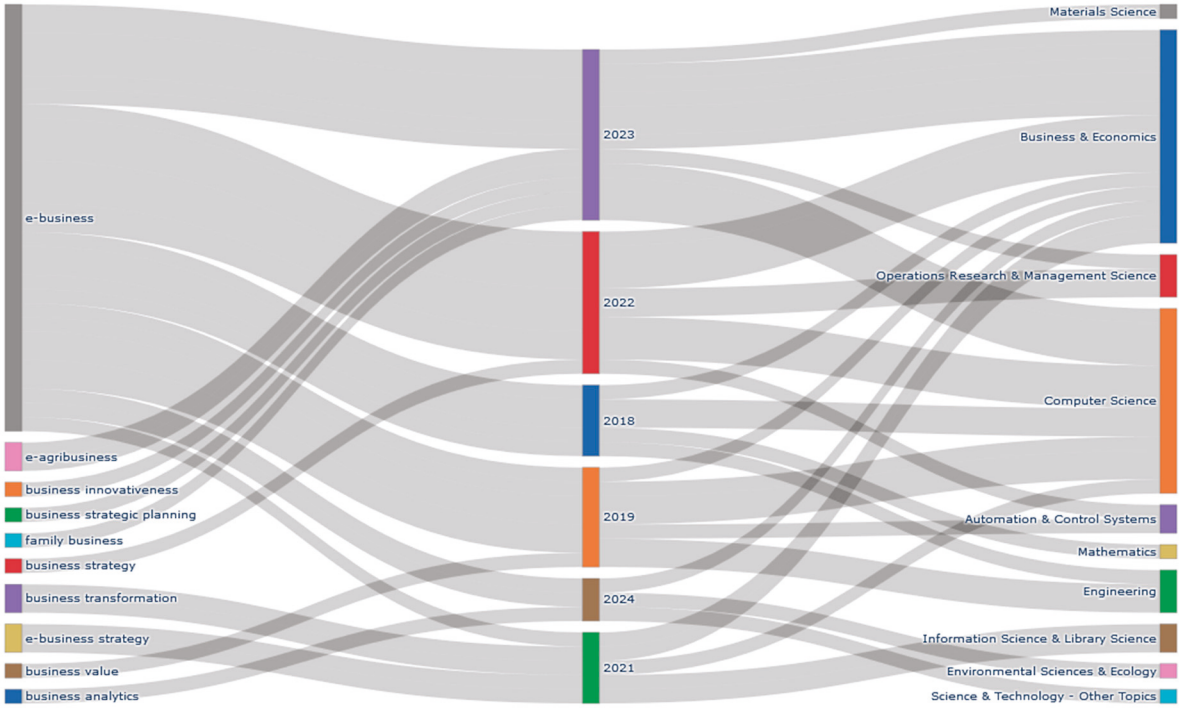


Figure 8. Sankey diagram linking "business" term to research areas and publications years.

4.2. Sentiment Analysis

The results of the sentiment analysis are briefly presented in Table 6.

Table 6. Sentiment analysis custom word lists, VADER, TextBlob libraries.

Year	Custom Pos	Custom Neg	Custom Neu	VADER Neg	VADER Neu	VADER Pos	VADER Compound	TextBlob Pol.	TextBlob Sub.
1999	0.00	0.00	93.00	0.00	0.87	0.13	0.89	0.09	0.39
2000	0.25	0.25	127.00	0.02	0.89	0.09	0.73	0.12	0.43
2001	0.13	0.38	122.50	0.03	0.84	0.13	0.85	0.12	0.52
2002	0.48	0.21	139.73	0.02	0.86	0.12	0.76	0.14	0.45
2003	1.00	0.57	168.29	0.03	0.86	0.11	0.65	0.10	0.54
2004	0.32	0.32	133.84	0.02	0.89	0.09	0.69	0.10	0.45
2005	0.48	0.33	128.00	0.03	0.88	0.08	0.44	0.07	0.53
2006	0.47	0.42	121.63	0.03	0.87	0.10	0.55	0.09	0.48
2007	0.35	0.12	114.76	0.02	0.86	0.13	0.78	0.12	0.47
2008	0.41	0.39	131.88	0.04	0.87	0.09	0.49	0.10	0.44
2009	0.50	0.25	136.75	0.02	0.85	0.12	0.86	0.06	0.53
2010	0.33	0.33	133.08	0.02	0.87	0.10	0.59	0.01	0.50
2011	0.09	0.27	134.91	0.03	0.85	0.12	0.83	0.11	0.47
2012	0.78	0.22	138.89	0.03	0.84	0.13	0.82	0.09	0.50

Table 6. *Cont.*

Year	Custom Pos	Custom Neg	Custom Neu	VADER Neg	VADER Neu	VADER Pos	VADER Compound	TextBlob Pol.	TextBlob Sub.
2013	0.50	0.10	149.80	0.02	0.88	0.11	0.69	0.07	0.50
2014	0.53	0.20	178.07	0.03	0.86	0.11	0.70	0.11	0.44
2015	0.36	0.36	128.73	0.03	0.87	0.09	0.64	0.05	0.38
2016	0.43	0.57	153.00	0.04	0.87	0.09	0.50	0.05	0.36
2017	0.57	0.36	155.71	0.03	0.88	0.09	0.57	0.12	0.46
2018	0.71	0.14	158.43	0.03	0.86	0.11	0.69	0.07	0.44
2019	0.68	0.21	159.74	0.03	0.84	0.13	0.76	0.08	0.44
2020	0.79	0.42	178.26	0.03	0.83	0.14	0.83	0.07	0.49
2021	0.80	0.10	246.40	0.02	0.87	0.11	0.82	0.05	0.48
2022	0.78	0.30	186.13	0.03	0.85	0.12	0.72	0.08	0.44
2023	0.56	0.24	210.08	0.03	0.84	0.12	0.77	0.05	0.45
2024	1.07	0.53	265.13	0.03	0.86	0.11	0.66	0.05	0.48

The sentiment analysis results using *custom word lists* reveal the progression of positive, negative and neutral sentiments over the years usingS a custom list of words. Positive sentiment shows fluctuations but demonstrates a general upward trend, with significant peaks observed in 2003, 2012, 2018 and 2021, and a notable high in 2024, reaching 1.07. Negative sentiment also fluctuates, maintaining relatively stable values in earlier years but peaking during specific periods, such as 2003 and 2016, indicating moments of heightened negativity. Neutral sentiment, on the other hand, exhibits a steady upward trend, particularly from 2013 onwards, culminating in its highest value of 265.13 in 2024, which reflects increasing neutrality in sentiment over time. Certain years stand out for their unique sentiment patterns. In 2003, there was a simultaneous spike in both positive and negative sentiments, suggesting a highly polarized year. In contrast, 2011 shows very low positive sentiment and relatively subdued negative sentiment, indicating a year of more balanced or less emotionally charged discourse. In recent years, especially from 2021 to 2024, there has been a strong increase in both positive and neutral sentiments, suggesting a shift towards more optimistic and balanced sentiment patterns. Throughout the dataset, neutral sentiment consistently dominates over positive and negative sentiments, which, as expected, indicates a preference for balanced or non-polarized language.

Moreover, as expected, neutral sentiment is the dominant category with VADER across all years, consistently above 83%, indicating a majority of neutral language. Positive sentiment remains relatively stable over time, typically ranging between 8% and 14%, with minor peaks such as 14% in 2020. Negative sentiment is the smallest category, generally below 4%, suggesting limited negative language. The compound score varies significantly, reflecting the overall sentiment polarity. Higher scores (e.g., 0.89 in 1999 and 0.86 in 2009) indicate strongly positive sentiment, while lower scores (e.g., 0.43 in 2005) suggest weaker positive sentiment or more mixed sentiment. Years such as 2020 and 2021 show high compound scores, indicating increasingly positive trends in sentiment during this period.

Sentiment analysis results using TextBlob, focusing on polarity (sentiment strength and direction) and subjectivity (degree of opinion versus fact) over time, are also presented in Table 5. Polarity values range between 0 and 1, where higher values indicate more positive sentiment. The data show fluctuations in polarity, with peaks in years like 2002 (0.14), 2001 (0.12) and 2007 (0.12). In contrast, years like 2016 (0.0457) and 2024 (0.0465) exhibit lower polarity, suggesting less positivity during these times. Subjectivity values, also ranging from 0 to 1, reflect how subjective or opinionated the content is. Higher subjectivity values

appear in years like 2003 (0.54) and 2005 (0.53), suggesting more opinion-based content. Conversely, lower subjectivity values, such as 2016 (0.35) and 1999 (0.38), indicate more objective content. While there is no consistent trend, subjectivity appears to peak and dip periodically.

While the three methods align in broad trends, with positive sentiment rising over time and neutral sentiment dominating, they differ in granularity and emphasis. Custom analysis provides more pronounced peaks and valleys, while VADER and TextBlob exhibit more stable, subdued trends.

4.3. Latent Dirichlet Allocation

It is a probabilistic generative model designed to uncover hidden topics in large text corpora. It operates on the assumption that documents in a dataset are composed of multiple latent topics, with each topic represented as a probability distribution over words. The primary objective of LDA is to identify these topic distributions, thereby unveiling the underlying thematic structure within the dataset. LDA's core principles revolve around a generative process and its grounding in Bayesian frameworks. The generative process assumes that each document is generated through a mixture of topics, where each topic is a probability distribution over a fixed vocabulary. For every word in a document, a topic is chosen based on the document's topic distribution, and a word is selected from the chosen topic's word distribution. The Bayesian framework introduces two critical parameters: alpha (α), which controls the sparsity of topic distributions across documents, and beta (β), which governs the sparsity of word distributions across topics. A high α results in documents covering many topics, while a high β implies that topics include a broad range of words. LDA employs optimization algorithms, such as Gibbs sampling or variational inference, to estimate the latent variables, including the topic distributions for each document and the word distributions for each topic.

The implementation of LDA begins with data preprocessing, where raw text is cleaned by removing stop words, punctuation and irrelevant terms. Techniques like tokenization, lowercasing and stemming/lemmatization standardize the text, converting the corpus into a document-term matrix representing the frequency of words in each document.

The model is initialized by defining the number of topics (k) and setting the hyper-parameters α and β, based on prior assumptions about sparsity. During training, LDA iteratively estimates the latent topic and word distributions. Each word in a document is probabilistically assigned to one of the k topics based on the current estimates of the document's topic distribution and the topic's word distribution. These assignments are refined over multiple iterations.

Evaluation of the LDA model is conducted using coherence scores, which assess the interpretability of the topics. A well-fitted model is indicated by low perplexity, while high coherence scores suggest meaningful and interpretable topics. Visualizations, often produced using tools like pyLDAvis, provide intertopic distance maps and term relevance plots, aiding in understanding topic relationships and key defining terms. When applied to research on e-business and artificial intelligence, LDA proves to be a powerful tool for identifying dominant research themes, emerging trends and essential focus areas. However, challenges persist, such as the need to predetermine the number of topics, which is often subjective and may require extensive tuning.

The LDA diagrams in Figure 9 provide a comprehensive overview of the dominant topics and their associated terms within the dataset related to e-business and artificial intelligence. The intertopic distance map, located on the left of each visualization, illustrates the spatial distribution of the identified topics in a two-dimensional space using multidimensional scaling. Larger circles represent topics with broader ranges of associated

terms, while the degree of separation or overlap indicates the level of uniqueness or shared relevance between the topics. The right side of each visualization highlights the top 30 most relevant terms for each topic. These terms are ranked by their estimated frequency within the topic and their overall saliency in the corpus. Terms shown in red specifically contribute to the selected topic, while terms in blue represent their overall frequency within the dataset.

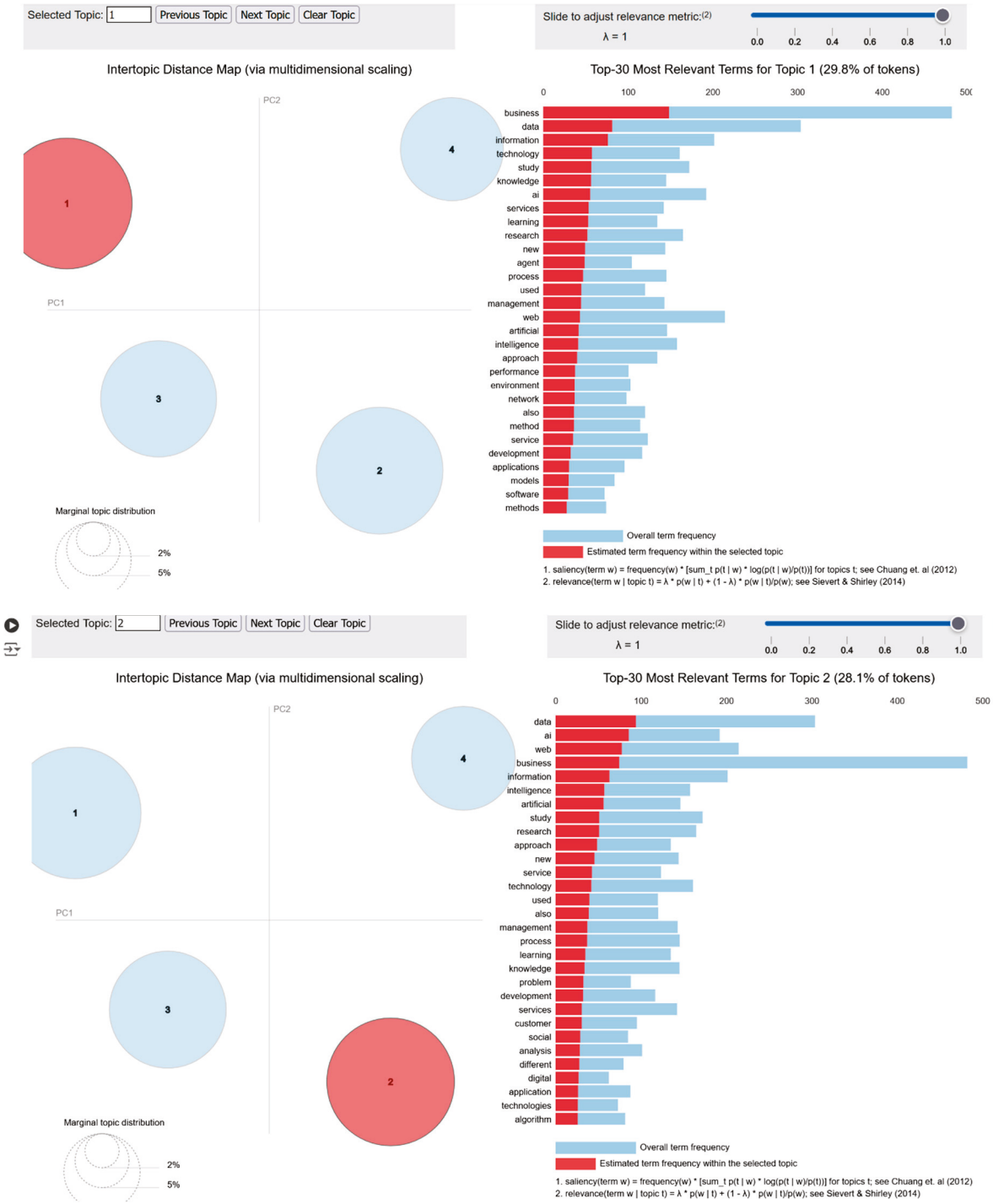


Figure 9. *Cont.*

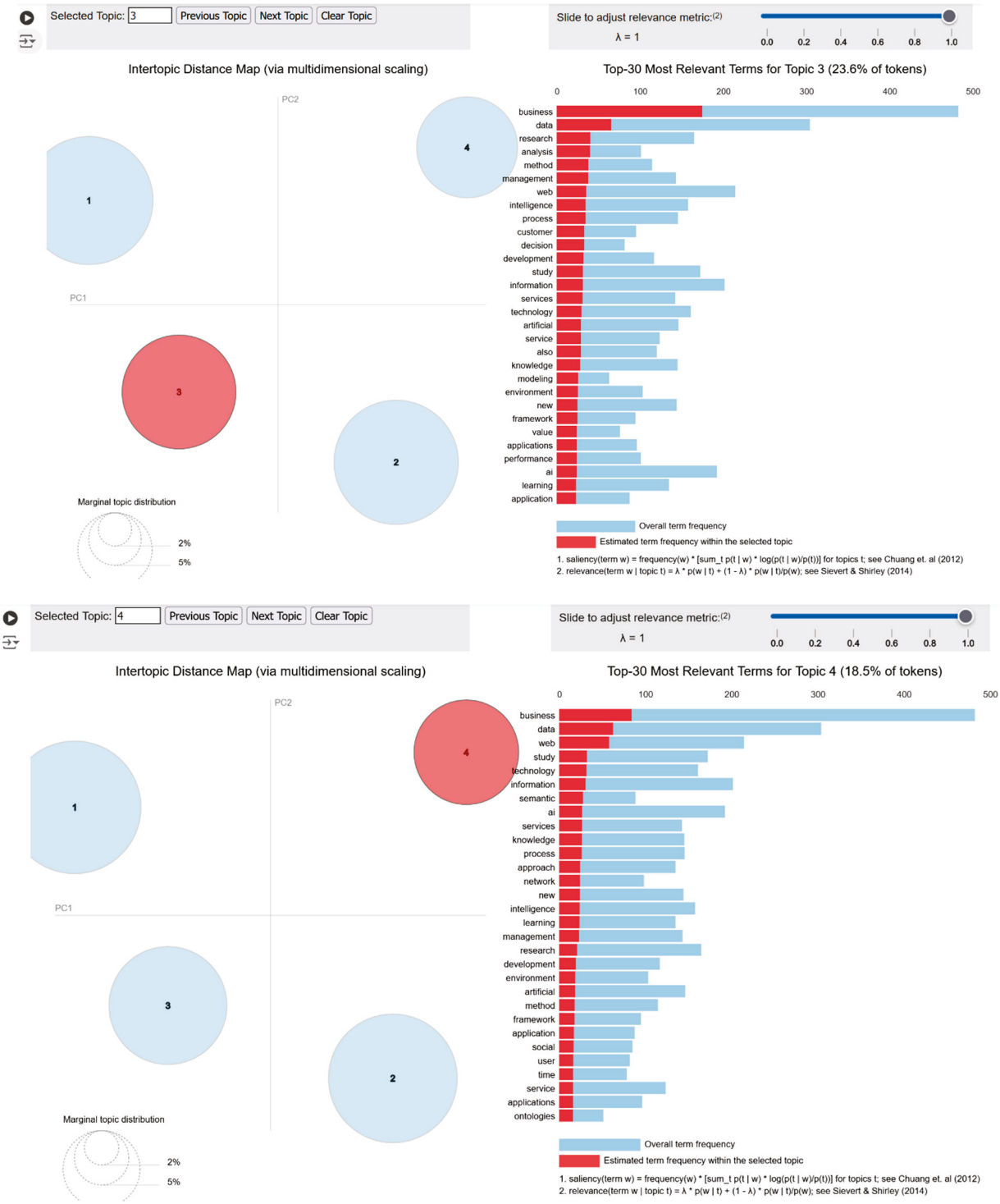


Figure 9. Topics graphical representation. For topics see Chuang et al., 2012 [45] and Sievert & Shirley, 2014 [46].

To answer the first research question (RQ1: What are the main topics embedded in the e-business and AI-related publications?), the topics identified with LDA are described in the following paragraphs.

Topic 1 (Data-Driven E-Business Strategies), which accounts for 29.8% of the tokens, emphasizes the intersection of data-driven decision-making and e-business frameworks. Key terms such as "business", "data", "information" and "technology" suggest a strong focus on how data systems, AI integration and information sharing enable business processes and service management.

Topic 2 (AI-Powered Web Solutions), representing 28.1% of the tokens, explores the role of AI and ML in web-based business applications. Terms such as "AI", "web", "business" and "intelligence" underscore the focus on leveraging AI to solve business challenges, enhancing customer service and innovating within web technologies. The prominence of terms like "service" and "technology" indicates a clear, practical application of AI in improving business workflows. *Topic 3 (Analytical Frameworks for Decision-Making)*, with 23.6% of the tokens, centers on analytical methods and decision-making frameworks within e-business. This topic is characterized by terms such as "analysis", "decision-making", "method" and "process", which point to the use of data modeling and research to optimize customer behavior analysis, business strategies and operational processes. The inclusion of terms like "environment" reflects an additional focus on adapting business processes to external conditions or sustainability considerations. *Topic 4 (Semantic Web and Intelligent Systems)*, which accounts for 18.5% of the tokens, delves into the semantic web and intelligent systems. Terms such as "semantic", "ontologies", "knowledge" and "web" indicate a research emphasis on organizing and interpreting data using semantic technologies to enhance intelligent systems. This topic reflects the role of advanced technologies in creating efficient knowledge systems and improving web-based platforms.

Across all four topics, certain themes are recurrent, such as the foundational role of data and the close connection between AI and business applications. While some topics, such as Topic 1 and Topic 2, share an emphasis on AI and data technologies, others, like Topic 3 and Topic 4, exhibit greater specialization in areas like process optimization and semantic systems. The model with four topics, asymmetric Alpha, and Beta = 0.91 provides the most coherent topics, suggesting this configuration best captures meaningful patterns in the corpus (as in Table 7). The model was tuned considering between 3 and 11 topics.

Table 7. Tuning the parameters of the probabilistic generative model.

Validation Set	Topics	Alpha	Beta	Coherence
75% Corpus	4	asymmetric	0.9099	0.6644
75% Corpus	5	0.9099	0.61	0.6206
75% Corpus	4	0.9099	0.9099	0.6164
75% Corpus	3	asymmetric	0.9099	0.6142
75% Corpus	5	asymmetric	0.9099	0.6037
75% Corpus	3	0.9099	0.9099	0.6005

5. Discussions on the Future of E-Business

The future of e-business powered by AI holds immense potential to redefine how businesses operate, interact with customers and innovate. AI advancements will drive new levels of efficiency, personalization and automation, creating experiences and opportunities that are currently unimaginable. To answer the second research question (RQ2: What are the key topics that could be further studied in the research landscape?), by complementarity with the LDA results, AI may elevate e-business in the future by the following:

(1) Hyper-personalized customer experiences. AI will refine personalization to a higher degree. Future e-businesses could offer real-time, context-aware recommendations based on detailed customer behavior, mood analysis and even biometric data. AI systems could predict customer needs before they are explicitly stated, offering products or services proactively. For example, smart assistants may analyze one's schedule and suggest purchases or services tailored to the upcoming activities.

(2) Voice and conversational commerce. Voice assistants and conversational AI will play a central role in e-business. Advanced NLP (speech-to-text) will enable seamless,

intuitive shopping experiences through voice commands. Customers may engage in multi-step, dynamic conversations to search for products, compare options and complete transactions. These assistants may also integrate with smart home devices to suggest purchases based on household needs.

(3) Immersive shopping with AR. AI will enhance AR to create immersive e-business environments. Customers could virtually "try on" clothes, visit hotel rooms, test furniture in their living spaces or explore virtual showrooms. AI algorithms may optimize these experiences by tailoring them to individual preferences, ensuring that the visualizations are accurate and engaging.

(4) Predictive and proactive commerce. AI will enable businesses to shift from reactive to proactive commerce. Predictive analytics powered by AI will identify what customers need before they realize it themselves. For instance, a grocery delivery service might automatically generate a shopping list based on previous purchases and predicted consumption patterns, delivering essentials before customers run out.

(5) Autonomous supply chains. Future supply chains will be fully autonomous, driven by AI and robotics. Predictive logistics will anticipate demand surges and optimize inventory and shipping routes dynamically. Autonomous drones and vehicles may handle last-mile delivery, ensuring faster and more efficient service. AI-powered systems will also improve supply chain resilience by predicting (e.g., based on local events in social media) and mitigating risks in real-time.

(6) Emotion-aware AI. AI systems of the future may become emotion-aware. They may assess the ethical implications of business decisions and adjust strategies to align with customer values. Emotion recognition technologies could analyze customer sentiment during interactions, customizing responses to create a more empathetic and satisfying experience.

(7) Blockchain integration for transparent transactions. AI and blockchain technologies will converge to enhance transparency and security in e-business. Smart contracts powered by AI could automate transaction verification and enforcement, reducing fraud and disputes. Blockchain-backed AI systems might also provide customers with full traceability of products, ensuring ethical sourcing and sustainability.

(8) Dynamic marketplaces. AI could revolutionize online marketplaces by creating dynamic, self-regulating platforms (e.g., for local electricity trade). These platforms would analyze real-time data to adjust pricing, match buyers with sellers more effectively and optimize product listings for visibility and sales. AI agents could even negotiate on behalf of buyers and sellers, creating a more efficient marketplace ecosystem.

(9) Advanced sentiment analysis and brand management. AI will become more adept at analyzing customer sentiment across multiple channels, including social media, reviews and direct feedback. Future systems will detect sentiment and also predict trends and public opinion shifts.

(10) Enhanced fraud prevention and security. AI will advance fraud detection by incorporating behavioral biometrics, anomaly detection and multi-layered authentication systems. Future e-business platforms may use AI to identify sophisticated cyber threats in real time, ensuring transactions remain secure. Blockchain and AI integration would also enhance trust and reduce fraud in online marketplaces.

6. Conclusions and Practical Implications

Our research seeks to reveal how AI-driven innovations are transforming e-business strategies, operations and sustainability practices. It identifies significant trends, such as the resurgence of interest in this research area post-2018, driven by factors like the COVID-19 pandemic, and highlights the dominant role of Chinese institutions in global collaborations. Furthermore, this research delves into the thematic prominence of keywords like "business

transformation", "e-business strategy" and the emerging "e-agribusiness", linking them to broader fields such as environmental sustainability. For instance, businesses may integrate AI-driven e-agribusiness solutions to achieve economic and environmental benefits by leveraging smart farming platforms. These systems optimize resource usage, enhance crop yields, and reduce costs through real-time data analysis and predictive analytics. Further, AI may enhance the traceability of products in e-agribusiness, benefiting both economic and environmental outcomes. By integrating AI with technologies such as blockchain, businesses create transparent and immutable records of a product's path from farm to consumer.

Through sentiment analysis, this study analyzes shifts in emotional tones within the literature, revealing a rise in positive sentiment over time. Lastly, the identification of four balanced topics underscores key areas of focus in the e-business and AI landscape: data-driven decision-making, AI in web-based business, analytical methods for decision-making and advancements in semantic web technologies. This research contributes to advancing knowledge in e-business and AI by mapping current trends and proposing future research directions.

The findings of this research have several practical implications across various domains:

- Enhanced business strategies. The emphasis on keywords like "business transformation", "business value" and "e-business strategy" highlights how AI and data-driven approaches can reshape traditional business models. For instance, Amazon uses AI-driven recommendations to transform customer purchasing experiences, generating significant revenue from cross-selling and up-selling. Similarly, AI-powered dynamic pricing models are employed by airlines and e-commerce platforms to adjust prices based on demand, competitor activity and customer behavior, enhancing revenue and customer satisfaction.
- Targeted investment in AI research. The surge in publications post-2018, particularly during the COVID-19 pandemic, indicates a growing reliance on AI to solve complex challenges. For example, AI-driven tools like Zoom and Microsoft Teams revolutionized remote work and education, enabling businesses and institutions to maintain operations during lockdowns. In healthcare, AI applications like telemedicine and predictive analytics are interesting achievements.
- Global collaboration opportunities. The dominance of Chinese institutions in the collaboration network presents opportunities for international partnerships. For instance, collaborations between Alibaba and European e-commerce businesses have leveraged AI to optimize supply chain logistics and improve market reach.
- Applications in sustainability. The connection between "e-agribusiness" and "Environmental Sciences & Ecology" underscores the role of AI and e-business in sustainable agriculture. For example, precision agriculture platforms use AI to analyze soil data and weather conditions, optimizing crop yields while reducing environmental impact. Additionally, AI-powered drones in India are used to monitor crop health and reduce pesticide usage, promoting eco-friendly practices.
- Improved customer experience. The findings related to web-based AI applications for customer service and workflow optimization indicate practical opportunities for businesses to enhance customer engagement, streamline operations and innovate in e-commerce platforms. Chatbots provide personalized product recommendations and answer queries in real-time, improving shopping experiences.
- Advancements in knowledge systems. The exploration of the semantic web and ontologies can drive the development of intelligent systems that organize and interpret vast amounts of data. These systems can be applied to improve search engines, recommendation systems and knowledge-sharing platforms. Google's Knowledge

Graph is an example, providing users with structured and meaningful search results. In certain e-learning platforms, AI organizes course content based on user preferences, enhancing learning experiences through personalized recommendations.

- Guidance for research and policy. The identification of latent themes and research gaps provides a roadmap for future studies in e-business and AI. For example, the rise of AI ethics committees in certain tech companies is shaping policies to address algorithmic fairness and data privacy. These efforts ensure responsible AI deployment while fostering public trust.
- AI-driven decision support. The focus on data-driven decision-making and analytical methods indicates practical pathways for implementing AI-powered tools in business environments. For example, AI-based tools like Tableau and Power BI enable real-time data visualization, helping managers make informed decisions. Predictive analytics platforms in retail, forecast inventory needs, optimize supply chains and reduce costs by ensuring products are available when and where they are needed.

The main limitations of the current research are (1) the coherence score of 0.66 indicates moderate interpretability of the identified topics. Higher scores might be achievable with alternative topic modeling techniques like BERTopic; (2) while this study mentions applications like "e-agribusiness", it does not explore sector-specific implications in depth; (3) this paper attributes a surge in publications around 2020 to the COVID-19 pandemic, but does not fully explore how pandemic-driven changes in research priorities and practices might have long-term implications for e-business and AI.

Two directions for future research are identified: (1) exploring alternative or hybrid topic modeling approaches could enhance the coherence and interpretability of the latent topics, potentially yielding more actionable insights; (2) investigating the role of emerging technologies, such as blockchain, quantum computing or generative AI, in transforming e-business could open new avenues for exploration.

Author Contributions: S.-V.O.: conceptualization, methodology, validation, formal analysis, investigation, resources, data curation, writing—original draft, writing—review and editing, visualization, supervision. A.B.: formal analysis, investigation, validation, writing—original draft, writing—review and editing, visualization, project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from the Ministry of Research, Innovation and Digitization, CNCS/CCCDI—UEFISCDI, project number COFUND-CETP-SMART-LEM-1, within PNCDI IV.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be available on request.

Conflicts of Interest: The authors declare that there are no conflicts of interest.

References

1. Huo, D.; Chen, Y.; Hung, K.; Ouyang, R.; Sun, B.; Cai, Y. Quality management of e-business: A key node analysis of ecological network in digital economy by using artificial intelligence. *Rom. J. Econ. Forecast.* **2019**, *22*, 166–179.
2. Purnomo, A.; Aziz, A.; Afia, N.; Sari, A.K.; Primadani, T.I.W. E-CRM: Three Decades of A Bibliometric Networks Visualizing of Academic Publication. In Proceedings of the 3rd International Conference on Cybernetics and Intelligent Systems, ICORIS 2021, Makasar, Indonesia, 25–26 October 2021. [CrossRef]
3. Tulcanaza-Prieto, A.B.; Cortez-Ordoñez, A.; Lee, C.W. Influence of Customer Perception Factors on AI-Enabled Customer Experience in the Ecuadorian Banking Environment. *Sustainability* **2023**, *15*, 12441. [CrossRef]
4. Arora, A.; Gupta, S.; Devi, C.; Walia, N. Customer experiences in the era of artificial intelligence (AI) in context to FinTech: A fuzzy AHP approach. *Benchmarking Int. J.* **2023**, *30*, 4342–4369. [CrossRef]

5. Nicolescu, L.; Tudorache, M.T. Human-Computer Interaction in Customer Service: The Experience with AI Chatbots—A Systematic Literature Review. *Electronics* **2022**, *11*, 1579. [CrossRef]
6. Huong, T.T.T.; Hanh, N.T.; Trang, H.T.D.; Chi, N.T.K. The impact of AI chatbots on customer experience in online retailing in an emerging economy. *Int. J. Process Manag. Benchmarking* **2023**, *15*, 182–197. [CrossRef]
7. Kanbach, D.K.; Heiduk, L.; Blueher, G.; Schreiter, M.; Lahmann, A. The GenAI is out of the bottle: Generative artificial intelligence from a business model innovation perspective. *Rev. Manag. Sci.* **2024**, *18*, 1189–1220. [CrossRef]
8. Sharma, A. Analyzing the Role of Artificial Intelligence in Predicting Customer Behavior and Personalizing the Shopping Experience in Ecommerce. *Int. J. Sci. Res. Eng. Manag.* **2023**, *7*, 1–8. [CrossRef]
9. Modgil, S.; Singh, R.K.; Hannibal, C. Artificial intelligence for supply chain resilience: Learning from COVID-19. *Int. J. Logist. Manag.* **2022**, *33*, 1246–1268. [CrossRef]
10. Sharma, R.; Shishodia, A.; Gunasekaran, A.; Min, H.; Munim, Z.H. The role of artificial intelligence in supply chain management: Mapping the territory. *Int. J. Prod. Res.* **2022**, *60*, 7527–7550. [CrossRef]
11. Oprea, S.-V.; Bâra, A. Detecting Malicious Uniform Resource Locators Using an Applied Intelligence Framework. *Comput. Mater. Contin.* **2024**, *79*, 3827–3853. [CrossRef]
12. Al Adwan, A.; Aladwan, R. Use of artificial intelligence system to predict consumers' behaviors. *Int. J. Data Netw. Sci.* **2022**, *6*, 1223–1232. [CrossRef]
13. Biswas, B.; Sanyal, M.K.; Mukherjee, T. AI-Based Sales Forecasting Model for Digital Marketing. *Int. J. E-bus. Res.* **2023**, *19*, 1–14. [CrossRef]
14. Vaduva, A.-G.; Oprea, S.-V.; Barbu, D.-C. Understanding Customers' Opinion using Web Scraping and Natural Language Processing. *Ovidius Univ. Ann. Econ. Sci. Ser.* **2023**, *1*, 537–544. [CrossRef]
15. Ratushniak, M. Business response to market tendencies: Consumerism, AI, e-commerce, and their influence on modern societies. *Bus. Econ. Sustain. Leadersh. Innov.* **2020**, *5*, 4–7. [CrossRef]
16. ThiDang, N.; MinhNguyen, T. Impact of Artificial Intelligence on E-Commerce Businesses in Ho Chi Minh City. *Glob. Acad. J. Econ. Bus.* **2022**, *4*, 146–151. [CrossRef]
17. Arman, M.; Lamiyar, U.R. Exploring the Implication of ChatGPT AI for Business: Efficiency and Challenges. *Int. J. Mark. Digit. Creat.* **2023**, *1*, 64–84. [CrossRef]
18. Yigitcanlar, T.; Degirmenci, K.; Inkinen, T. Drivers behind the public perception of artificial intelligence: Insights from major Australian cities. *AI Soc.* **2024**, *39*, 833–853. [CrossRef] [PubMed]
19. Lari, H.A.; Vaishnava, K.; Manu, K.S. Artifical Intelligence in E-commerce: Applications, Implications and Challenges. *Asian J. Manag.* **2022**, *13*, 235–244. [CrossRef]
20. Bansal, C.; Pandey, K.K.; Goel, R.; Sharma, A.; Jangirala, S. Artificial intelligence (AI) bias impacts: Classification framework for effective mitigation. *Issues Inf. Syst.* **2023**, *24*, 367–389. [CrossRef]
21. Janssen, M.; Hartog, M.; Matheus, R.; Ding, A.Y.; Kuk, G. Will Algorithms Blind People? The Effect of Explainable AI and Decision-Makers' Experience on AI-supported Decision-Making in Government. *Soc. Sci. Comput. Rev.* **2022**, *40*, 478–493. [CrossRef]
22. Areiqat, A.Y.; Alheet, A.F.; Qawasmeh, R.A.A.; Zamil, A.M. Artificial intelligence and its drastic impact on e-commerce progress. *Acad. Strateg. Manag. J.* **2021**, *20*, 1–11.
23. Wang, C.; Ahmad, S.F.; Ayassrah, A.Y.B.A.; Awwad, E.M.; Irshad, M.; Ali, Y.A.; Al-Razgan, M.; Khan, Y.; Han, H. An empirical evaluation of technology acceptance model for Artificial Intelligence in E-commerce. *Heliyon* **2023**, *9*, e18349. [CrossRef] [PubMed]
24. Thamik, H.; Wu, J. The Impact of Artificial Intelligence on Sustainable Development in Electronic Markets. *Sustainability* **2022**, *14*, 3568. [CrossRef]
25. Arulprakash, E.; Martin, A. An object-oriented neural representation and its implication towards explainable AI. *Int. J. Inf. Technol.* **2024**, *16*, 1303–1318. [CrossRef]
26. Al Wahshat, H.; Abu-Ulbeh, W.; Yusoff, M.H.; Zakaria, M.D.; Hamzah, W.M.A.F.W.; Stenin, N.P. The Detection of E-Commerce Manipulated Reviews Using GPT-4. In Proceedings of the 2023 International Conference on Computer Science and Emerging Technologies, CSET 2023, Bangalore, India, 10–12 October 2023. [CrossRef]
27. Bawack, R.E.; Wamba, S.F.; Carillo, K.D.A.; Akter, S. Artificial intelligence in E-Commerce: A bibliometric study and literature review. *Electron. Mark.* **2022**, *32*, 297–338. [CrossRef] [PubMed]
28. Goti, A.; Querejeta-Lomas, L.; Almeida, A.; de la Puerta, J.G.; López-de-Ipiña, D. Artificial Intelligence in Business-to-Customer Fashion Retail: A Literature Review. *Mathematics* **2023**, *11*, 2943. [CrossRef]
29. Thayyib, P.V.; Mamilla, R.; Khan, M.; Fatima, H.; Asim, M.; Anwar, I.; Shamsudheen, M.K.; Khan, M.A. State-of-the-Art of Artificial Intelligence and Big Data Analytics Reviews in Five Different Domains: A Bibliometric Summary. *Sustainability* **2023**, *15*, 4026. [CrossRef]
30. Espina-Romero, L.; Guerrero-Alcedo, J. Fields Touched by Digitalization: Analysis of Scientific Activity in Scopus. *Sustainability* **2022**, *14*, 14425. [CrossRef]

31. Rydell, L. Predictive Algorithms, Data Visualization Tools, and Artificial Neural Networks in the Retail Metaverse. *Linguist. Philos. Investig.* **2022**, *21*, 25–40. [CrossRef]
32. Frioui, S.; Graa, A. Bibliometric Analysis of Artificial Intelligence in the Scope of E-Commerce: Trends and Progress over the Last Decade. *Manag. Econ. Rev.* **2024**, *9*, 5–24. [CrossRef]
33. Erdoğan, U. A systematic review on the use of artificial intelligence in e-commerce. *Toplum Ekon. ve Yönetim Derg.* **2023**, *4*, 184–197. [CrossRef]
34. Widjaja, W.; Rustam, Y.W.A. E-Retailing study: A bibliometric analysis and systematic literature review. *TEKNOSAINS J. Sains Teknol. Dan Inform.* **2024**, *11*, 112–119. [CrossRef]
35. Desai, P.; Ganatra, K. Artificial Intelligence In Strengthening The Operations of Ecommerce Based Business. In Proceedings of the 2022 International Conference on Interdisciplinary Research in Technology and Management, IRTM 2022—Proceedings, Kolkata, India, 24–26 February 2022. [CrossRef]
36. Louis, J.V.; Noerlina; Syahchari, D.H. Digital business transformation: Analysis of the effect artificial intelligence in e-commerce's product recommendation. *Adv. Inf. Syst.* **2024**, *8*, 64–69. [CrossRef]
37. Zhang, D.; Pee, L.G.; Cui, L. Artificial intelligence in E-commerce fulfillment: A case study of resource orchestration at Alibaba's Smart Warehouse. *Int. J. Inf. Manag.* **2021**, *57*, 102304. [CrossRef]
38. Pallathadka, H.; Ramirez-Asis, E.H.; Loli-Poma, T.P.; Kaliyaperumal, K.; Ventayen, R.J.M.; Naved, M. Applications of artificial intelligence in business management, e-commerce and finance. *Mater. Today Proc.* **2023**, *80*, 2610–2613. [CrossRef]
39. Ziakis, C.; Vlachopoulou, M. Artificial Intelligence in Digital Marketing: Insights from a Comprehensive Review. *Information* **2023**, *14*, 664. [CrossRef]
40. Qin, M.; Zhu, W.; Zhao, S.; Zhao, Y. Is Artificial Intelligence Better than Manpower? The Effects of Different Types of Online Customer Services on Customer Purchase Intentions. *Sustainability* **2022**, *14*, 3974. [CrossRef]
41. Rajeshwari, S.; Praveenadevi, D.; Revathy, S.; Sreekala, S.P. Utilizing AI technologies to enhance e-commerce business operations. In *Toward Artificial General Intelligence: Deep Learning, Neural Networks, Generative AI*; De Gruyter: Berlin, Germany, 2023. [CrossRef]
42. Von Zahn, M.; Feuerriegel, S.; Kuehl, N. The Cost of Fairness in AI: Evidence from E-Commerce. *Bus. Inf. Syst. Eng.* **2022**, *64*, 335–348. [CrossRef]
43. Khrais, L.T. Role of artificial intelligence in shaping consumer demand in e-commerce. *Future Internet* **2020**, *12*, 226. [CrossRef]
44. Fatima, S. AI Implementation in an E-commerce. *Int. J. Multidiscip. Res.* **2023**, *5*, 1–11. [CrossRef]
45. Chuang, J.; Manning, C.D.; Heer, J. Termite: Visualization techniques for assessing textual topic models. In Proceedings of the International Working Conference on Advanced Visual Interfaces, Capri Island, Italy, 21–25 May 2012; pp. 74–77.
46. Sievert, C.; Shirley, K. LDAvis: A method for Visualizing and Interpreting Topics. In Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, Baltimore, MD, USA, 27 June 2014; pp. 63–70.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

A Tripartite Evolutionary Game-Based Cooperation Model of Cross-Border E-Commerce Logistics Alliances: A Case Study of China

Xiaohong Miao [1,2], Zhongbin Li [1,*], Yingzheng Yan [1] and Anxin Xu [3]

[1] College of Economics and Management, Fujian Agriculture and Forestry University, Fuzhou 350002, China; miaoxiaohong369852@163.com (X.M.); yingz_yan012@163.com (Y.Y.)
[2] College of Business Administration, Fujian Business University, Fuzhou 350012, China
[3] School of Journalism and Communication, Minjiang University, Fuzhou 350108, China; anxin_xu2011@163.com
* Correspondence: zhongbin_li0922@163.com; Tel.: +86-139-0590-5427

Abstract: As a new business model, cross-border e-commerce has become an important way for countries to meet new foreign trade requirements in the Internet economy. The cross-border logistics industry plays a crucial role in supporting cross-border e-commerce. Compared with domestic e-commerce, cross-border logistics faces more challenges. To address the problems in cross-border logistics, this study takes China as an example and constructs a tripartite evolutionary game model to facilitate information collaboration among cross-border e-commerce platforms, domestic logistics enterprises, and foreign logistics enterprises. The collaboration strategies in this tripartite information system are simulated using MATLAB. The study highlights key factors affecting information cooperation, such as standardization levels, risk and payoff distributions, and their implications on collaboration decisions. Specifically, the results show that higher levels of information collaboration standardization promote cooperative strategies among players; the risk associated with information collaboration is the most sensitive factor influencing decision-making within cross-border logistics alliances; and when the payoff distribution coefficient is too high, other members may resist cooperation.

Keywords: supply chain cooperation; simulation; cross-border e-commerce; evolutionary game theory; sustainable development

1. Introduction

With the rapid development of the digital economy and the advancement of China's "Belt and Road" Initiative, the cross-border e-commerce sector in China has reached a critical growth phase. According to the China Cross-Border E-Commerce Market Data Report 2023, China's cross-border e-commerce market reached CNY 2.38 trillion in the same year, exhibiting a 15.6% increase compared to the previous year. There are differences in labor costs between countries, leading to new profit drivers. In addition, the general decline in tariff levels has improved the investment climate and international trade, and capital and other factors of production have become globally mobile [1]. These factors have lengthened the production chain for products and led to the widespread distribution of different production factors across different countries, with the sources of product value involving many countries or regions. The development of cross-border e-commerce drives the globalization of the industry while also affecting the upstream manufacturing of goods and the downstream consumption of Chinese goods [2]. At the same time, foreign demand

is growing, but cross-border logistics still faces significant shortcomings [3]. The diversity, small batch size, and high frequency of cross-border e-commerce commodities necessitate the development of cross-border logistics in China. However, cross-border logistics enterprises are currently underdeveloped and face challenges related to information sharing, reliability, multi-party cooperation, product quality, and operational costs. These issues hinder the efficiency and integration of cross-border e-commerce logistics. To address these challenges, this study explores information collaboration challenges in cross-border e-commerce logistics alliances, aiming to enhance logistics efficiency and support the sustainable development of China's cross-border e-commerce industry.

2. Literature Review

Research on cooperation in cross-border e-commerce logistics has been conducted using both qualitative and quantitative approaches, each offering valuable insights into various aspects of logistics collaboration. However, a clear synthesis is necessary to understand how these studies contribute to the evolving role of information collaboration in improving cross-border logistics efficiency. This review organizes the existing research into four key themes: (1) theoretical frameworks for cooperation, (2) operational challenges in cross-border logistics, (3) collaborative decision-making and payoff distribution, and (4) information collaboration in logistics alliances. By structuring the discussion around these themes, this review identifies existing gaps and establishes the foundation for this study's focus on the dynamic evolution of information collaboration.

2.1. Theoretical Frameworks for Cooperation

A number of studies have explored theoretical models for cooperation in cross-border e-commerce logistics. Togar proposed a supply chain cooperation integration framework based on a peer-to-peer approach, which is valuable for enhancing cooperation between logistics and e-commerce platforms [4–6]. While Togar's framework provides a solid foundation for understanding supply chain integration, it primarily focuses on domestic logistics and does not adequately address the unique challenges of cross-border logistics, such as the complexities of international regulations, customs clearance, and the need for real-time information sharing across borders. Similarly, Tenorio developed a cooperation framework based on an analysis of supply chain mechanisms, emphasizing the importance of effective collaboration for operational efficiency [7]. However, these studies focus largely on theoretical frameworks and do not directly address the challenges specific to cross-border logistics, such as high operational costs and information tracking issues, which are critical in the context of e-commerce. In examining the integration of cross-border logistics and e-commerce, Ye and Zhang highlighted the interdependent relationship between the two and suggested that integration will become more pronounced in the future [8,9]. However, their study lacks a detailed exploration of how information collaboration can facilitate this integration.

2.2. Operational Challenges in Cross-Border Logistics

Cross-border logistics face several operational challenges, including high costs, long delivery cycles, and difficulties in information tracking. Wu identified a lack of coordination between cross-border logistics and e-commerce, noting challenges such as high operational costs, long cycle times, and difficulty in tracking information [10]. While these findings provide valuable insights, they do not offer concrete solutions for improving information collaboration. Similarly, Yang echoed these concerns, but did not discuss how cooperation frameworks could address these operational challenges, especially in terms of improving logistics efficiency [11]. These studies highlight the need for more research on

how information collaboration can mitigate these challenges, particularly through the use of advanced technologies and standardized processes.

2.3. Collaborative Decision-Making and Payoff Distribution

Quantitative studies have focused on collaborative decision-making and payoff distribution methods in supply chains, offering valuable insights into the stability and efficiency of logistics alliances. Zhang developed a Nash equilibrium model under symmetric information, providing equilibrium solutions for logistics service providers and integrators in cross-border e-commerce. The study considered the provider's ability to clear customs and the influence of parameters such as the customer's penalty for the integrator in coordinating supply chain quality [12]. While this model offers important perspectives on the coordination of logistics in cross-border e-commerce, it overlooks the dynamic nature of information collaboration and does not address the evolving risks and benefits associated with long-term cooperation in logistics alliances. Similarly, Gurbanova used a Bayesian game model to analyze the payoff game behavior of alliance members and showed how each player can strategize to maximize the overall payoff [13]. However, this analysis is limited in that it does not consider the impact of asymmetric information on collaboration or explore the potential for free-riding behavior, a critical issue in cross-border logistics.

Zhang further emphasized that a payoff-sharing agreement could increase the motivation of alliance members under specific conditions [14]. While this finding is important for motivating cooperation, it does not delve into the complexities of information sharing or explore how variations in information reliability and transparency might affect member cooperation. Du et al. combined system dynamics with an evolutionary game model to simulate a 4PL cross-border logistics alliance under different strategies, concluding that platform supervision, logistics enterprise efforts, and merchant participation would eventually reach equilibrium [15]. This model is valuable for understanding stability, but it does not address the evolving dynamics of information collaboration over time, a key aspect of cross-border logistics alliances. Wang's study considered consumer satisfaction and proposed a game model that included cross-border suppliers, cross-border e-commerce enterprises, retailers, and consumers, investigating the supply chain's revenue and cost-sharing contracts under both centralized and decentralized decision-making conditions [16]. While the study emphasizes cost-sharing and revenue distribution, it overlooks the role of information collaboration and how sharing information across parties could enhance overall performance and reduce risks associated with coordination.

2.4. Information Collaboration in Cross-Border Logistics Alliances

Various studies have specifically investigated the issue of information collaboration in cross-border e-commerce alliances, which is crucial for improving the efficiency and sustainability of these alliances. For example, Barut et al. measured the participation of supply chain enterprises in information collaboration, using the degree of information extension and the strength of information as collaboration metrics [17]. While this study provides insight into how information sharing can be quantified, it lacks a focus on the evolving dynamics and interdependencies between logistics partners in cross-border e-commerce. Similarly, Cui et al. employed a fuzzy planning model to explore the optimal path for a 4PL logistics alliance, accounting for factors such as the overall abilities of logistics enterprises and cost discounts [18]. This model highlights the importance of strategic planning, but it overlooks the critical role of information exchange in enhancing cooperation among cross-border logistics partners.

Dai et al. utilized an evolutionary game model, combined with system dynamics simulation, to investigate the stability of logistics alliances and found that the scale of logistics

resource investment and organizational costs can enhance cooperation [19]. While these findings contribute to our understanding of resource allocation in logistics alliances, they do not address the specific challenges associated with information collaboration in the context of cross-border e-commerce. Chen examined the impact of factors like information technology stability, data validity, and operational security on information collaboration behavior among logistics enterprises [20]. These factors are key to ensuring successful collaboration, yet the study does not explore how they interact or evolve within cross-border logistics alliances. Zhou et al. summarized the motivations for innovation in cross-border e-commerce operation modes, proposing a "TaaS" operation mode based on information collaboration [21]. This framework is insightful but lacks an exploration of how information collaboration dynamics can change over time as the market evolves. Liu et al. identified the problems in cross-border e-commerce information collaboration and proposed a blockchain-based model to enhance it [22]. While this model is promising for ensuring transparency, it does not consider the strategic decision-making processes involved in information sharing among logistics partners. Some studies also analyzed the synergistic mechanisms between cross-border e-commerce and logistics based on supply chain theory, emphasizing that the amount of information collaboration in logistics impacts the efficiency of logistics resource allocation. Rashid and Rasheed investigated the impact of logistics service quality dimensions on product satisfaction in the e-commerce industry of Pakistan and found that product quality and information quality had a significant positive impact on product satisfaction, while the impacts of product availability, delivery time, and shipping cost were insignificant. This research provides a new perspective on the study of logistics service-related factors from the perspective of consumer satisfaction [23]. While their work highlights the importance of information collaboration, it does not delve into the dynamic and evolving nature of these interactions, which is central to the challenges faced by cross-border logistics alliances.

2.5. Identified Gaps and Research Contributions

Prior research has primarily addressed qualitative aspects such as development status, model analysis, and policy recommendations. Quantitative studies, while fewer in number, primarily concentrate on synergistic evaluation. However, there is a noticeable lack of research on the dynamic evolution of information collaboration in cross-border logistics alliances. This gap is particularly significant given the costs involved in information collaboration and the potential for 'free-riding', where one participant gains more benefit at the expense of others. To better understand the existing literature, Table 1 summarizes the key aspects covered by previous studies, including theoretical frameworks, operational challenges, collaborative decision-making, information collaboration, dynamic analysis, and risks. As shown in Table 1, while some studies have addressed theoretical frameworks and operational challenges, there is limited focus on the dynamic evolution of information collaboration and the associated risks, such as "free-riding". These gaps highlight the need for more comprehensive research on the dynamic and evolving nature of information collaboration in cross-border logistics alliances.

This study aims to address these gaps by combining theoretical frameworks with empirical evidence to explore the cooperative dynamics of cross-border e-commerce logistics alliances. Specifically, it identifies the key factors affecting information collaboration stability and proposes a cross-border logistics alliance model based on information collaboration. Using evolutionary game theory, the study analyzes the trajectory and evolutionary stability strategies in the information collaboration behavior of the three players, considering the effects of information standardization level, information collaboration risk, and information collaboration payoff distribution. Finally, MATLAB is used to simulate and reveal the mechanism of information collaboration in cross-border logistics alliances. This research offers

valuable insights for the stable development of cross-border logistics alliances and provides practical references for the development of cross-border e-commerce logistics in China. It also contributes to the improvement of related information collaboration systems, addressing the risks and dynamics that have been largely overlooked in previous studies. By integrating theoretical analysis with empirical simulation, this study not only fills critical gaps in the existing literature but also provides a robust framework for future research in this domain.

Table 1. Summary of the related literature regarding cooperation in cross-border e-commerce logistics.

Literature	Theoretical Frameworks	Operational Challenges	Collaborative Decision-Making	Information Collaboration	Dynamic Analysis	Risks
Togar [4–6], Tenorio [7]	√		√			
Ye and Zhang [8,9]	√			√		
Wu [10], Yang [11]		√				√
Zhang [12], Gurbanova [13], Zhang [14], Du et al. [15], Wang [16]		√	√	√		
Barut et al. [17], Cui et al. [18], Dai et al. [19], Chen [20], Zhou et al. [21], Liu et al. [22], Xie et al. [23]	√	√		√		
Rashid and Rasheed [24]	√	√				
Our Study	√	√		√	√	√

3. Evolutionary Game Model Description and Assumptions

3.1. Game Players and Cooperation Strategies

The process of building a cross-border e-commerce logistics alliance for information collaboration involves dynamic interactions among multiple parties. Hence, this research aims to establish a tripartite evolutionary game model encompassing cross-border e-commerce platforms, foreign logistics enterprises, and domestic logistics enterprises, as shown in Figure 1. All participants in the game exhibit bounded rationality as they seek to optimize their own interests and iteratively adapt their strategies throughout the game.

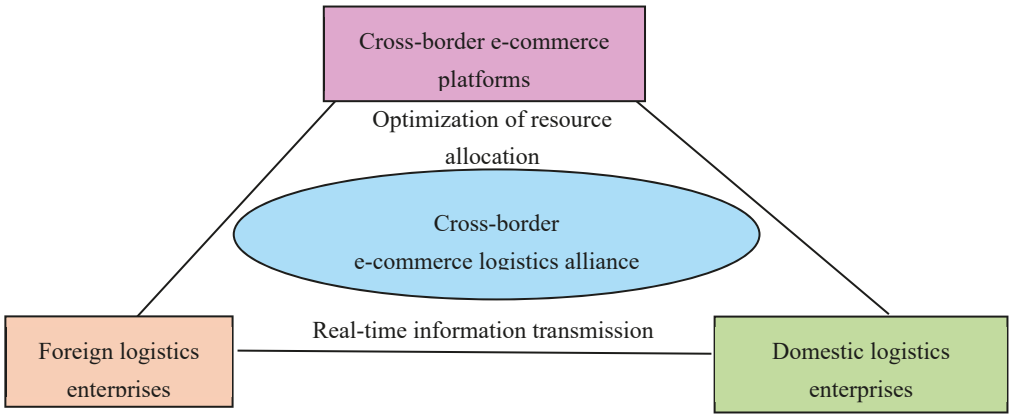


Figure 1. Dynamic interactions and collaborative information decisions among three parties in cross-border logistics alliances.

According to actual needs, the alliance members can choose to collaborate on information, which is manifested in each member's commitment to establishing an information collaboration platform to achieve information system integration, real-time information transmission and sharing, and value co-creation [24]. The three players can also choose

not to collaborate on information, in which case each player will independently operate their own information system, with simple information exchange. Thus, the set of choice strategies in the game is {cooperation, non-cooperation}.

3.2. Assumptions

Based on the analysis of the key factors affecting information collaboration in cross-border logistics alliances, the following four core assumptions are defined:

Assumption 1. Information Collaboration Standard Level: The level of information collaboration is measured by the integration and interaction of cross-border information systems. This assumption is based on the understanding that in a globalized business environment, efficient information sharing and integration among different countries and regions are crucial for the success of cross-border logistics alliances. The integration of information systems allows for seamless communication and data exchange, reducing misunderstandings and inefficiencies. Let A represent the collaboration standard, β be the total information, and μ_i (for i = 1,2,3) be the information conversion coefficient for each player. The direct payoff for each player is proportional to the collaboration level, with higher standards resulting in greater information and higher payoffs. This is because as the level of information collaboration increases, players can access more accurate and timely information, which enables them to make better decisions and gain more benefits. If only two players collaborate, the payoff depends on the shared information, denoted as δ. This reflects the fact that in a partial collaboration scenario, the benefits are still related to the quality and quantity of shared information.

Assumption 2. Information Collaboration Risk Cost: When the three players engage in information collaboration, the costs incurred by the cross-border e-commerce platform, domestic logistics enterprises, and foreign logistics enterprises due to strategic, technological, and human-factor risks are denoted as C_1, C_2, and C_3, respectively, with $C_i > 0$. Higher collaboration risks result in higher costs due to uncertainty. In the context of cross-border logistics, where different countries and regions have diverse regulations and cultures, there are numerous potential risks that can affect the success of information collaboration. Strategic risks may include changes in trade policies, technological risks may involve data security and system failures, and human-factor risks may include misunderstandings and communication barriers. These risks introduce uncertainties into the collaboration process, leading to increased costs. For example, if there is a data breach due to a technological failure, the enterprises may face significant financial losses and reputational damage.

Assumption 3. Information Collaboration Payoff Distribution: The original income payoffs for each player are denoted as ω_i, and the additional collaboration payoff is π, which is the incremental benefit gained through cooperative efforts. The distribution coefficients ∂_i (where $\partial_1 + \partial_2 + \partial_3 = 1$) are used to allocate the cooperation payoff among the players. This assumption is based on the principle of fairness and efficiency in collaborative settings, ensuring that the total payoff is distributed in a reasonable manner. In cases of two-player collaboration, the payoff is distributed based on coefficients α_i; this reflects the different contributions or power dynamics between the two collaborating players. The non-collaborating player incurs an economic loss σ_i, which serves as a disincentive for choosing not to participate in the information collaboration process. For example, if a player decides not to collaborate while the others do, they may miss out on potential benefits and instead face financial losses.

Assumption 4. Information Collaboration Input Cost: The total cost of labor, materials, and technology development for information collaboration is denoted as F; this cost includes the resources expended by the alliance members to establish and maintain the information collaboration system. These costs are shared among the alliance members to

ensure a collective responsibility and investment in the collaborative effort. The individual contributions of each member are represented as F_i, which allows for a detailed accounting of each member's input to the information collaboration process. This breakdown helps in evaluating the fairness of the cost sharing and identifying areas where individual members may need to adjust their contributions for better overall performance. For instance, if a member's F_i is relatively high compared to others, it may indicate a need for the optimization or a reallocation of resources within the alliance.

3.3. Establishing the Revenue Matrix

Assume that the probability of cross-border e-commerce platforms choosing information collaboration is x, and the probability of choosing non-cooperation is $1 - x$. Similarly, the probability of foreign logistics enterprises choosing information collaboration is y, and the probability of choosing non-cooperation is $1 - y$. The probability of domestic logistics enterprises choosing information collaboration is z, and the probability of choosing non-cooperation is $1 - z$, where $x, y, z \in [0, 1]$.

To provide clarity on the variables used in the analysis, Table 2 summarizes the definitions and descriptions of all key parameters.

Table 2. Variable definitions and descriptions.

Variable	Description	Range/Constraints
x	Probability of cross-border e-commerce platforms choosing information collaboration	$x \in [0, 1]$
y	Probability of foreign logistics enterprises choosing information collaboration	$y \in [0, 1]$
z	Probability of domestic logistics enterprises choosing information collaboration	$z \in [0, 1]$
ω_i	Original income payoff for enterprise i (i = 1,2,3)	$\omega_i \geq 0$
A	Level of information collaboration standard	$A \geq 0$
μ_i	Information conversion coefficient for enterprise i	$\mu_i \geq 0$
β	Total amount of information when all three players collaborate	$\beta \geq 0$
δ	Total amount of information when only two players collaborate	$\delta \geq 0$
$A\mu_i\beta$	Combined effect of information collaboration level A, conversion coefficient μ_i, and total information β	$A\mu_i\beta \geq 0$
$A\mu_i\delta$	Combined effect of collaboration level A, conversion coefficient μ_i, and information δ when only two players collaborate	$A\mu_i\delta \geq 0$
π	Additional payoff from collaboration when all three players collaborate	$\pi \geq 0$
∂_i	Payoff distribution coefficient for enterprise i when all three collaborate	$\partial_i \in [0, 1]$, $\partial_1 + \partial_2 + \partial_3 = 1$
D	Additional payoff from collaboration when only two players collaborate	$D \geq 0$
α_i	Payoff distribution coefficient for enterprise i when only two collaborate	$\alpha_i \in [0, 1]$, $\alpha_1 + \alpha_2 = 1$, $\alpha_2 + \alpha_3 = 1$, or $\alpha_1 + \alpha_3 = 1$
C_i	Risk cost incurred by enterprise i due to information collaboration	$C_i \geq 0$
F_i	Input cost borne by enterprise i for information collaboration	$F_i \geq 0$
σ_i	Economic loss incurred by enterprise i if it chooses non-cooperation while the other two collaborate	$\sigma_i \geq 0$

Based on the above assumptions and parameters, a payoff matrix for the tripartite evolutionary game is constructed, as shown in Table 3.

Table 3. Payoff matrix for the tripartite evolutionary game.

Cross-Border e-Commerce Platforms	Foreign Logistics Enterprises	Domestic Logistics Enterprises	
		Cooperation (z)	**Non-Cooperation ($1-z$)**
Cooperation (x)	Cooperation (y)	$\omega_1 + A\mu_1\beta + \pi\partial_1 - C_1 - F_1$ $\omega_2 + A\mu_2\beta + \pi\partial_2 - C_2 - F_2$ $\omega_3 + A\mu_3\beta + \pi\partial_3 - C_3 - F_3$	$\omega_1 + A\mu_1\delta + D\alpha_1 - C_1 - F_1$ $\omega_2 + A\mu_2\delta + D\alpha_2 - C_2 - F_2$ $\omega_3 - \sigma_3$
	Non-cooperation ($1-y$)	$\omega_1 + A\mu_1\delta + D\alpha_1 - C_1 - F_1$ $\omega_2 - \sigma_2$ $\omega_3 + A\mu_3\delta + D\alpha_3 - C_3 - F_3$	$\omega_1 - C_1$ ω_2 ω_3
Non-cooperation ($1-x$)	Cooperation (y)	$\omega_1 - \sigma_1$ $\omega_2 + A\mu_2\delta + D\alpha_2 - C_2 - F_2$ $\omega_3 + A\mu_3\delta + D\alpha_3 - C_3 - F_3$	ω_1 $\omega_2 - C_2$ ω_3
	Non-cooperation ($1-y$)	ω_1 ω_2 $\omega_3 - C_3$	ω_1 ω_2 ω_3

4. Solving the Evolutionary Stabilization Strategy

4.1. Return Expectation Function Construction

Cross-border e-commerce platforms, foreign logistics enterprises, and domestic logistics enterprises can all choose the best strategy based on their actual payoffs and risks [25]. The expected returns for cross-border e-commerce platforms choosing information collaboration or information non-cooperation are denoted as $\forall_{x1}$ and $\forall_{x2}$, respectively, while the average return is denoted as $\forall_x$; these can be obtained as follows:

$$\begin{aligned}\forall_{x1} = yz(\omega_1 + A\mu_1\beta + \pi\partial_1 - C_1 - F_1) + y(1-z)(\omega_1 + A\mu_1\delta + D\alpha_1 - C_1 - F_1)\\ +z(1-y)(\omega_1 + A\mu_1\delta + D\alpha_1 - C_1 - F_1) + (1-z)(1-y)(\omega_1 - C_1),\end{aligned} \tag{1}$$

$$\forall_{x2} = yz(\omega_1 - \sigma_1) + y(1-z)\omega_1 + z(1-y)\omega_1 + (1-z)(1-y)\omega_1 = \omega_1 - yz\sigma_1, \tag{2}$$

$$\forall_x = x\forall_{x1} + (1-x)\forall_{x2}. \tag{3}$$

Equation (1) calculates the expected return for cross-border e-commerce platforms when they choose to collaborate. It considers the payoffs from collaboration with both foreign and domestic logistics enterprises, as well as the costs and risks involved. Equation (2) calculates the expected return for cross-border e-commerce platforms when they choose not to collaborate. It accounts for the potential loss (σ_1) due to non-cooperation. Equation (3) represents the weighted average of the expected returns from collaboration and non-cooperation, based on the probability x of choosing collaboration.

Similarly, the expected returns for foreign logistics firms choosing information collaboration or information non-cooperation strategies are denoted as $\forall_{y1}$ and $\forall_{y2}$, respectively, while the average expected return is denoted as $\forall_y$, which can be obtained as follows:

$$\begin{aligned}\forall_{y1} = xz(\omega_2 + A\mu_2\beta + \pi\partial_2 - C_2 - F_2) + x(1-z)(\omega_2 + A\mu_2\delta + D\alpha_2 - C_2 - F_2)\\ +z(1-x)(\omega_2 + A\mu_2\delta + D\alpha_2 - C_2 - F_2) + (1-z)(1-x)(\omega_2 - C_2),\end{aligned} \tag{4}$$

$$\forall_{y2} = xz(\omega_2 - \sigma_2) + x(1-z)\omega_2 + z(1-x)\omega_2 + (1-z)(1-x)\omega_2 = \omega_2 - xz\sigma_2, \tag{5}$$

$$\forall_y = y\forall_{y1} + (1-y)\forall_{y2}. \tag{6}$$

Similarly, the expected returns for domestic logistics enterprises choosing information collaboration or information non-cooperation strategies are denoted as $\forall_{z1}$ and $\forall_{z2}$,

respectively, while the average expected return is denoted as $\forall_z$, which can be obtained as follows:

$$\begin{aligned}\forall_{z1} = xy(\omega_3 + \mathrm{A}\mu_3\beta + \pi\partial_3 - C_3 - F_3) + x(1-y)(\omega_3 + \mathrm{A}\mu_3\delta + D\alpha_3 - C_3 - F_3)\\ +y(1-x)(\omega_3 + \mathrm{A}\mu_3\delta + D\alpha_3 - C_3 - F_3) + (1-x)(1-y)(\omega_3 - C_3),\end{aligned} \tag{7}$$

$$\forall_{z2} = xy(\omega_3 - \sigma_3) + x(1-y)\omega_3 + y(1-x)\omega_3 + (1-x)(1-y)\omega_3 = \omega_3 - xy\sigma_3, \tag{8}$$

$$\forall_{\mathrm{z}} = \mathrm{z}\forall_{\mathrm{z1}} + (1-\mathrm{z})\forall_{\mathrm{z2}}. \tag{9}$$

4.2. Analysis of the Evolutionary Paths of the Three Players

4.2.1. Evolutionary Path of Cross-Border E-Commerce Platforms

Based on the above analysis, the replication dynamic equation for the information collaboration strategy of cross-border e-commerce platforms is the following:

$$\begin{aligned}F(x) = \tfrac{dx}{dt} = x(1-x)(\forall_{x1} - \forall_{x2})\\ = x(1-x)[yz(\omega_1 + A\mu_1\beta + \pi\partial_1 - C_1 - F_1) + \omega_1 + y(1-z)(\omega_1 + A\mu_1\delta + D\alpha_1 - C_1 - F_1)\\ +z(1-y)(\omega_1 + A\mu_1\delta + D\alpha_1 - C_1 - F_1) + (1-z)(1-y)(\omega_1 - C_1) - (\omega_1 - yz\sigma_1)]\\ = x(1-x)[yz(A\mu_1\beta - 2A\mu_1\delta + \pi\partial_1 - 2D\alpha_1 + F_1 + \sigma_1) + (y+z)(A\mu_1\delta + D\alpha_1 - F_1) - C_1].\end{aligned} \tag{10}$$

Equation (10) describes the rate of change in the proportion of cross-border e-commerce platforms choosing collaboration. It depends on the difference between the expected returns of collaboration and non-cooperation.

To facilitate analysis, let z_0 be the point where $F(x) = 0$, as derived from the stability theorem of the replicated dynamic equations. This can be obtained as follows:

$$z_0 = \frac{C_1 - y(A\mu_1\delta + D\alpha_1 - F_1)}{y(A\mu_1\beta - 2A\mu_1\delta + \pi\partial_1 - 2D\alpha_1 + F_1 + \sigma_1) + (A\mu_1\delta + D\alpha_1 - F_1)}. \tag{11}$$

(1) If $z = z_0$, then $F(x) \equiv 0$. This implies that all levels are in a steady state; that is, cross-border e-commerce platforms choosing information collaboration is a stable strategy regardless of the percentage, and it will not change over time.

(2) If $z \neq z_0$, such that $F(x) = 0$, two equilibrium points, x = 0 and x = 1, can be obtained. Consider the two points in the following two cases:

① When $z_0 < z < 1$, $\frac{dF(x)}{dt}\mid_{x=0} > 0$, and $\frac{dF(x)}{dt}\mid_{x=1} < 0$, x = 1 is the evolutionary stability point. That is, when the proportion of domestic logistics enterprises choosing information collaboration exceeds z_0, cross-border e-commerce platforms will ultimately choose information collaboration.

② When $0 < z < z_0$, $\frac{dF(x)}{dt}\mid_{x=0} < 0$, and $\frac{dF(x)}{dt}\mid_{x=1} > 0$, x = 0 is the evolutionary stability point. That is, when the proportion of domestic logistics enterprises choosing information collaboration is lower than z_0, cross-border e-commerce platforms will ultimately choose information non-cooperation.

Based on the above analysis, we can see that the evolution of cross-border e-commerce platforms' collaborative strategy selection is influenced not only by the strategy selection of domestic logistics enterprises but also by that of foreign logistics enterprises.

4.2.2. Evolutionary Path of Foreign Logistics Enterprises

Based on the above analysis, the replication dynamic equation for the information collaboration strategy of foreign logistics enterprises can be expressed as follows:

$$\begin{aligned}F(y) = \tfrac{dy}{dt} = y(1-y)(\forall_{y1} - \forall_{y2})\\ = y(1-y)[xz(\omega_2 + A\mu_2\beta + \pi\partial_2 - C_2 - F_2) + x(1-z)\omega_2 + A\mu_2\delta + D\alpha_2 - C_2 - F_2)\\ +z(1-x)\ (\omega_2 + A\mu_2\delta + D\alpha_2 - C_2 - F_2) + (1-z)(1-x)(\omega_2 - C_2) - (\omega_2 - xz\sigma_2)]\\ = y(1-y)[xz(A\mu_2\beta - 2A\mu_2\delta + \pi\partial_2 - 2D\alpha_2 + F_2 + \sigma_2) + (x+z)(A\mu_2\delta + D\alpha_2 - F_2) - C_2].\end{aligned} \tag{12}$$

Equation (12) describes the rate of change in the proportion of foreign logistics enterprises choosing collaboration, based on the difference between the expected returns of collaboration and non-cooperation.

To facilitate analysis, let x_0 be the point where $F(y) = 0$. From the stability theorem of the replicated dynamic equations, it can be expressed as follows:

$$x_0 = \frac{C_2 - z(A\mu_2\delta + D\alpha_2 - F_2)}{z(A\mu_2\beta - 2A\mu_2\delta + \pi\partial_2 - 2D\alpha_2 + F_2 + \sigma_2) + (A\mu_2\delta + D\alpha_2 - F_2)} \tag{13}$$

(1) If $x = x_0$, then $F(y) \equiv 0$. This implies that, at all levels, the strategy of foreign logistics enterprises choosing information collaboration is stable. It remains stable regardless of the proportion and does not change over time.

(2) If $x \neq x_0$, let $F(y) = 0$. Two equilibrium points, y = 0 and y = 1, can be obtained. Consider these two points in the following two cases:

① When $x_0 < x < 1$, $\frac{dF(y)}{dt}\mid_{y=0} > 0$, and $\frac{dF(y)}{dt}\mid_{y=1} < 0$. At this point, $y = 1$ becomes the evolutionary stability point; that is, when the proportion of cross-border e-commerce platforms choosing information collaboration exceeds x_0, they will ultimately choose information collaboration.

② When $0 < x < x_0$, $\frac{dF(y)}{dt}\mid_{y=0} < 0$, and $\frac{dF(y)}{dt}\mid_{y=1} > 0$. At this point, $y = 0$ becomes the evolutionary stability point; that is, when the proportion of cross-border e-commerce platforms choosing information collaboration is below x_0, they will ultimately choose information non-cooperation.

Based on the above analysis, it can be seen that the evolution of the collaborative strategy selection by foreign logistics enterprises is influenced not only by the strategy choices of cross-border e-commerce platforms but also by those of domestic logistics enterprises.

4.2.3. Evolutionary Path of Domestic Logistics Enterprises

Based on the above analysis, the replication dynamic equation for the information collaboration strategy of foreign logistics enterprises can be derived as follows:

$$\begin{aligned} F(z) = \tfrac{dz}{dt} = z(1-z)(\forall_{y1} - \forall_{y2}) \\ = z(1-z)[xy(\omega_3 + A\mu_3\beta + \pi\partial_3 - C_3 - F_3) + x(1-y)(\omega_3 + A\mu_3\delta + D\alpha_3 - C_3 - F_3) \\ + y(1-x)\,(\omega_3 + A\mu_3\delta + D\alpha_3 - C_3 - F_3) + (1-x)(1-y)(\omega_3 - C_3) - (\omega_3 - xz\sigma_3)] \\ = z(1-z)[xy(A\mu_3\beta - 2A\mu_3\delta + \pi\partial_3 - 2D\alpha_3 + F_3 + \sigma_3) + (x+y)(A\mu_3\delta + D\alpha_3 - F_3) - C_3]. \end{aligned} \tag{14}$$

Equation (14) describes the rate of change in the proportion of domestic logistics enterprises choosing collaboration, based on the difference between the expected returns of collaboration and non-cooperation.

To facilitate analysis, let y_0 be the point at which $F(z) = 0$. According to the stability theorem of the replicated dynamic equations, this can be derived as follows:

$$y_0 = \frac{C_3 - x(A\mu_3\delta + D\alpha_3 - F_3)}{x(A\mu_3\beta - 2A\mu_3\delta + \pi\partial_3 - 2D\alpha_3 + F_3 + \sigma_3) + (A\mu_3\delta + D\alpha_3 - F_3)} \tag{15}$$

(1) If $y = y_0$, then $F(z) \equiv 0$. This implies that at all levels, the strategy of foreign logistics enterprises choosing information collaboration is stable. It remains unchanged regardless of the proportion and does not fluctuate over time.

(2) If $y \neq y_0$, then $F(z) = 0$. Two equilibrium points, $z = 0$ and $z = 1$, can be found. Consider the following two cases for these points:

① When $y_0 < y < 1$, $\frac{dF(z)}{dt}\mid_{z=0} > 0$, and $\frac{dF(z)}{dt}\mid_{z=1} < 0$. At this point, $z = 1$ is the evolutionary stability point; that is, when the proportion of foreign logistics enterprises choosing information collaboration is higher than y_0, they will ultimately choose information collaboration.

② When $0 < y < y_0$, $\frac{dF(y)}{dt}\mid_{z=0} < 0$, and $\frac{dF(y)}{dt}\mid_{z=1} > 0$. At this point, $z = 0$ is the evolutionary stability point; that is, when the proportion of foreign logistics enterprises choosing information collaboration is lower than y_0, they will ultimately choose information non-cooperation.

Based on the above analysis, it can be seen that the evolution of domestic logistics enterprises' collaborative strategy selection is influenced not only by the strategy choices of cross-border e-commerce platforms but also by the strategy choices of domestic logistics enterprises themselves.

4.3. Analysis of the Evolutionary Stabilization Strategy of Tripartite Game

The evolutionary stability strategy of the system of differential equations can be derived from the Jacobi local stability analysis of this system [26]. The Jacobi matrix for the system can be computed as follows:

$$\begin{pmatrix} \begin{array}{l}(1-2x)[yz(A\mu_1\beta-2A\mu_1\delta\\ +\pi\partial_1-2D\alpha_1+F_1+\sigma_1)\\ +(y+z)(A\mu_1\delta+D\alpha_1-F_1)-C_1]\end{array} & \begin{array}{l}x(1-x)[z(A\mu_1\beta-2A\mu_1\delta\\ +\pi\partial_1-2D\alpha_1+F_1+\sigma_1)\\ +A\mu_1\delta+D\alpha_1-F_1]\end{array} & \begin{array}{l}x(1-x)[y(A\mu_1\beta-2A\mu_1\delta\\ +\pi\partial_1-2D\alpha_1+F_1+\sigma_1)\\ +A\mu_1\delta+D\alpha_1-F_1]\end{array} \\ \begin{array}{l}y(1-y)[z(A\mu_2\beta-2A\mu_2\delta\\ +\pi\partial_2-2D\alpha_2+F_2+\sigma_2)\\ +A\mu_2\delta+D\alpha_2-F_2]\end{array} & \begin{array}{l}(1-2y)[xz(A\mu_2\beta-2A\mu_2\delta\\ +\pi\partial_2-2D\alpha_2+F_2+\sigma_2)\\ +(x+z)(A\mu_2\delta+D\alpha_2-F_2)-C_2]\end{array} & \begin{array}{l}y(1-y)[x(A\mu_2\beta-2A\mu_2\delta\\ +\pi\partial_2-2D\alpha_2+F_2+\sigma_2)\\ +A\mu_2\delta+D\alpha_2-F_2]\end{array} \\ \begin{array}{l}z(1-z)[y(A\mu_3\beta-2A\mu_3\delta\\ +\pi\partial_3-2D\alpha_3+F_3+\sigma_3)\\ +A\mu_3\delta+D\alpha_3-F_3]\end{array} & \begin{array}{l}z(1-z)[x(A\mu_3\beta-2A\mu_3\delta\\ +\pi\partial_3-2D\alpha_3+F_3+\sigma_3)\\ +A\mu_3\delta+D\alpha_3-F_3]\end{array} & \begin{array}{l}(1-2z)[xy(A\mu_3\beta-2A\mu_3\delta\\ +\pi\partial_3-2D\alpha_3+F_3+\sigma_3)\\ +(x+y)(A\mu_3\delta+D\alpha_3-F_3)-C_3]\end{array} \end{pmatrix} \tag{16}$$

Using the Lyapunov indirect method, if all eigenvalues of the Jacobi matrix have negative real parts, the equilibrium point is asymptotically stable. If at least one eigenvalue has a positive real part, the equilibrium point is unstable. If all eigenvalues of the Jacobi matrix, except for those with zero real parts, have negative real parts, the equilibrium point is in a critical state, and its stability cannot be determined solely by the sign of the eigenvalues. Table 4 presents the stability analysis of each equilibrium point.

Table 4. Eigenvalues of the matrix.

Equilibrium Point	Eigenvalue
$E_1(0,0,0)$	$-C_1, -C_2, -C_3$
$E_2(1,0,0)$	$C_1, A\mu_2\delta+D\alpha_2-F_2-C_2, A\mu_3\delta+D\alpha_3-F_3-C_3$
$E_3(0,1,0)$	$A\mu_1\delta+D\alpha_1-F_1-C_1, C_2, A\mu_3\delta+D\alpha_3-F_3-C_3$
$E_4(0,0,1)$	$A\mu_1\delta+D\alpha_1-F_1-C_1, A\mu_2\delta+D\alpha_2-F_2-C_2, C_3$
$E_5(1,1,0)$	$F_1+C_1-A\mu_1\delta-D\alpha_1, F_2+C_2-A\mu_2\delta-D\alpha_2,$ $A\mu_3\beta+\pi\partial_3+\sigma_3-F_3-C_3$
$E_6(1,0,1)$	$F_1+C_1-A\mu_1\delta-D\alpha_1, A\mu_2\beta+\pi\partial_2+\sigma_2-F_2-C_2, F_3+C_3-A\mu_3\delta-D\alpha_3$
$E_7(0,1,1)$	$A\mu_1\beta+\pi\partial_1+\sigma_1-F_1-C_1, F_2+C_2-A\mu_2\delta-D\alpha_2, F_3+C_3-A\mu_3\delta-D\alpha_3$
$E_8(1,1,1)$	$-(A\mu_1\beta+\pi\partial_1+\sigma_1-F_1-C_1), -(A\mu_2\beta+\pi\partial_2+\sigma_2-F_2-C_2),$ $-(A\mu_3\beta+\pi\partial_3+\sigma_3-F_3-C_3)$

4.4. Analysis of Stabilization Conditions

It is assumed that the payoffs for the three players choosing information collaboration are greater than the payoffs for those choosing information non-cooperation—that is, as shown in Table 3,

$$\omega_i + \mathrm{A}\mu_i\beta + \pi\partial_i - C_i - F_i > \omega_i. \tag{17}$$

Therefore, for the results in Table 4, $\mathrm{A}\mu_i\beta + \pi\partial_i - C_i - F_i > 0$, so $-(A\mu_i\beta + \pi\partial_i + \sigma_i - F_i - C_i) < 0$.

Now, the evolutionary stability strategies under the two hypotheses are considered separately, as follows:

Assumption 1: When any two players choose information collaboration, the payoff from information collaboration is greater than that from information non-cooperation; that is,

$$\omega_i + \mathrm{A}\mu_i\delta + D\alpha_i - C_i - F_i > \omega_i. \tag{18}$$

Therefore, the results in Table 4, $\mathrm{A}\mu_i\delta + D\alpha_i - C_i - F_i > 0$, and the eigenvalues corresponding to the equilibrium points $E_1(0,0,0)$ and $E_8(1,1,1)$ are both less than 0. Thus, E_1 and E_8 are the stability points of the replicated dynamic systems, and their corresponding evolutionary stability strategies are both cooperative and non-cooperative. On the other hand, the eigenvalues of $E_2(1,0,0)$, $E_3(0,1,0)$, and $E_4(0,0,1)$ are all greater than 0. Therefore, they are saddle points, and the remaining points are unstable points.

Assumption 2: When any two players choose information collaboration, the payoff from information collaboration is less than the payoff from information non-cooperation; that is,

$$\omega_i + \mathrm{A}\mu_i\delta + D\alpha_i - C_i - F_i < \omega_i. \tag{19}$$

Therefore, the results in Table 4, $\mathrm{A}\mu_i\delta + D\alpha_i - C_i - F_i < 0$, and the eigenvalues corresponding to the equilibrium points $E_1(0,0,0)$ and $E_8(1,1,1)$ are both less than 0. Thus, E_1 and E_8 are the stabilization points of the replicated dynamic systems, and their corresponding evolutionary stabilization strategies are both cooperative and non-cooperative. Other than that, the eigenvalues of $E_5(1,1,0)$, $E_6(1,0,1)$, and $E_7(0,1,1)$ are all greater than 0. Thus, they are saddle points, and the rest of the points are unstable points.

Table 5 presents the analysis of the stability of equilibrium points in the tripartite information collaboration evolutionary game of cross-border logistics alliances.

Table 5. Equilibrium point stability analysis.

	Assumption A		**Assumption B**	
Equilibrium Point	**Sign**	**Stabilization Conclusion**	**Sign**	**Stabilization Conclusion**
$E_1(0,0,0)$	$(-,-,-)$	stability	$(-,-,-)$	stability
$E_2(1,0,0)$	$(+,+,+)$	saddle points	$(+,-,-)$	instability
$E_3(0,1,0)$	$(+,+,+)$	saddle points	$(-,+,-)$	instability
$E_4(0,0,1)$	$(+,+,+)$	saddle points	$(-,-,+)$	instability
$E_5(1,1,0)$	$(-,-,+)$	instability	$(+,+,+)$	saddle points
$E_6(1,0,1)$	$(-,+,-)$	instability	$(+,+,+)$	saddle points
$E_7(0,1,1)$	$(+,-,-)$	instability	$(+,+,+)$	saddle points
$E_8(1,1,1)$	$(-,-,-)$	stability	$(-,-,-)$	stability

5. Data Simulation Analysis

The rapid development of cross-border e-commerce platforms accelerates the transformation and upgrading of traditional logistics enterprises [27]. However, many challenges remain. Some logistics enterprises are conservative in their thinking, lack suitable management talent, and find transformation difficult. To address these issues, 10 senior managers from cross-border logistics and 15 system simulation experts were invited to discuss the parameter-setting problems. Based on this, the conditions that the initial parameters should meet are determined as follows:

(1) If the information standard is fully aligned, A = 1; if the information standard is not aligned, A = 0; and if the information standard is partially aligned, A $\in$ (0, 1).

(2) The payoff from tripartite information collaboration is greater than the payoff from tripartite information non-cooperation; thus, $A\mu_i\beta + \pi\partial_i - C_i - F_i > 0$.

(3) When any two players choose information collaboration, the normal payoff of the alliance members is higher than the payoff of information collaboration; thus, $A\mu_i\delta + D\alpha_i - C_i - F_i < 0$.

(4) Cross-border e-commerce platforms have more advanced technology and higher abilities to convert information into payoffs compared to domestic and foreign logistics enterprises [28]. However, the cross-border logistics businesses of domestic logistics enterprises are still in their infancy, with technology levels far below those of foreign logistics enterprises, and their information conversion factor is lower than that of foreign logistics enterprises. Therefore, $\mu_1 > \mu_2 > \mu_3$.

(5) Cross-border e-commerce platforms integrate the information resources of domestic and foreign logistics enterprises, coordinate the operations of the entire cross-border logistics alliance, and contribute the most to the overall business. Therefore, $\partial_1 > \partial_2$, and $\partial_1 > \partial_3$.

(6) The total amount of tripartite information collaboration and its additional payoffs are higher than those of any two-party information collaboration; that is, $\pi > D$.

(7) The cross-border e-commerce platforms are the dominant enterprises, and they bear the costs of information collaboration input [29]. Their economic losses from information non-cooperation should be higher than those of the other two enterprises. Cross-border e-commerce platforms possess more information and face a higher cost associated with the risk of information collaboration compared to the other two enterprises [30]. Moreover, cross-border e-commerce platforms have a greater payoff than both domestic and foreign logistics enterprises. Thus, $C_1 > C_2, C_1 > C_3$, $\sigma_1 > \sigma_2, \sigma_1 > \sigma_3$, $F_1 > F_2, F_1 > F_3$, $\omega_1 > \omega_2$, and $\omega_1 > \omega_3$.

The above conditions are synthesized, and the initial parameter values are set as shown in Table 6.

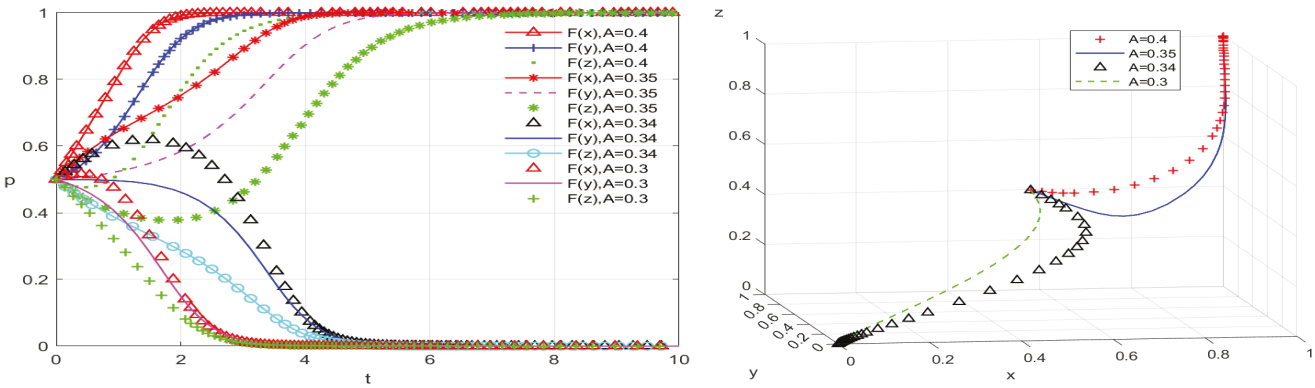
Table 6. Initial parameter value settings.

A	β	μ_1	μ_2	μ_3	π	∂_1	∂_2	∂_3	ω_1	ω_2	ω_3	α_i	D	σ_1	σ_2	σ_3	C_1	C_2	C_3	δ	F_1	F_2	F_3
0.5	30	0.5	0.3	0.2	6	0.4	0.3	0.3	8	5	5	0.5	4	2.5	1.5	1.5	3	2.5	2.5	20	2	1.5	1.5

Let the initial values of the tripartite participants' selection strategies (x, y, z) be (0.5, 0.5, 0.5). The three elements, A, ∂_i, and F_i, which have the greatest influence on the tripartite synergistic strategy of the cross-border logistics alliance, are selected for analysis. MATLAB 6.0 is used for the simulation.

5.1. Influence of Information Collaboration Standardization Level on Evolution Results

Figure 2 shows the simulation of the evolutionary trajectory of the tripartite information collaboration strategy of the cross-border logistics alliance under different levels of information collaboration standardization.



(**a**) (**b**)

Figure 2. Effect of the information collaboration criteria on the evolutionary results. (**a**) Two-dimensional evolutionary paths. (**b**) Three-dimensional evolutionary paths.

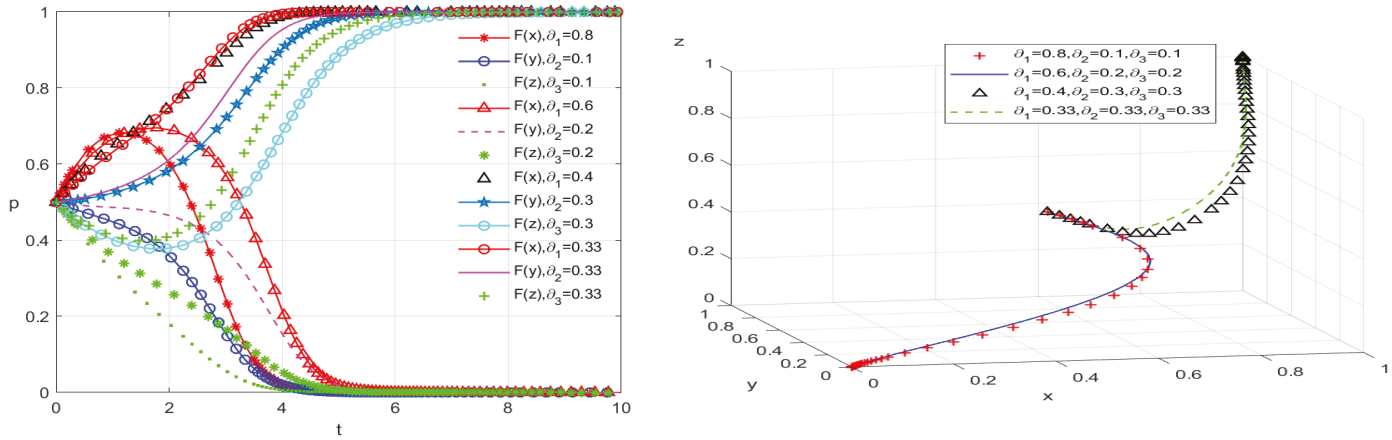
As shown in Figure 2a, the critical threshold for the information collaboration standardization level A lies between 0.34 and 0.35. When A is below this threshold, the evolutionary stable strategy for the tripartite information collaboration tends to be (0, 0, 0), indicating that no collaboration occurs. However, when A exceeds this critical value, the probabilities x, y, and z (representing the likelihood of cross-border e-commerce platforms, foreign logistics enterprises, and domestic logistics enterprises choosing information collaboration, respectively) converge to 1, signaling that all players are willing to collaborate on information. Notably, x converges to 1 faster than y and z. This is because cross-border e-commerce platforms typically have more advanced information technology and a stronger ability to convert information into revenue, which makes the payoff from information collaboration more substantial for them. As a result, they are more eager to adopt information collaboration strategies as the standardization level increases.

On the other hand, z, representing the domestic logistics enterprises' willingness to collaborate, converges to 1 more slowly. This slower convergence suggests that domestic logistics enterprises may be more cautious or face greater barriers (e.g., technological constraints, less-developed infrastructure, or less familiarity with information collaboration) in adopting these strategies. Once the other two players' willingness to collaborate reaches a certain threshold, the domestic logistics enterprises' cooperation willingness begins to rise significantly.

Figure 2b presents the three-dimensional evolutionary paths, where the strategies evolve from (0, 0, 0) to (1, 1, 1) as A increases, eventually stabilizing at a point of full collaboration. The results demonstrate that the level of information collaboration standards significantly influences the strategies of alliance members. Only when A > 0.34 do all three players opt for information collaboration. Moreover, as the level of information collaboration standards rises, the willingness of alliance members to collaborate strengthens. This is attributed to the standardization of technical standards for information transmission, the seamless transfer of cross-border electronic documents, the real-time sharing and transmission of cross-border information, and the uniformity of logistics-related terminology. These factors reduce misunderstandings and inefficiencies, thereby enhancing the economic performance of alliance enterprises.

5.2. Effect of Information Collaboration Payoff Distribution on Evolutionary Outcomes

Figure 3 illustrates the simulation of the evolution of cross-border logistics alliance members' information collaboration strategies under different information collaboration revenue allocation coefficients.



(**a**) (**b**)

Figure 3. Effect of information collaboration payoff distribution on evolutionary outcomes. (**a**) Two-dimensional evolutionary paths. (**b**) Three-dimensional evolutionary paths.

In this study, three distinct information collaboration payoff distribution combinations are analyzed: highly unreasonable payoff distribution (0.8, 0.1, 0.1), average payoff distribution (0.33, 0.33, 0.33), and moderate payoff distribution (0.4, 0.3, 0.3). As shown in Figure 3a, first, when the information collaboration payoff distribution is highly unreasonable, x initially converges rapidly to 1 but is eventually forced to decline as y and z quickly converge to 0, causing the equilibrium point to trend towards (0, 0, 0). This outcome occurs because the cross-border e-commerce platform's larger share of the payoff leads to a disproportionate benefit, reducing the willingness of the domestic and foreign logistics enterprises to cooperate. As a result, the entire system moves toward non-cooperation. Second, under the average payoff distribution, x, y, and z converge to 1 at the fastest rate. This indicates that when the differences in the risks associated with information collaboration and the strength of input investments between alliance members are minimal, a more equitable payoff distribution facilitates a quicker consensus on cooperation. In this case, the balanced distribution encourages all players to adopt the information collaboration strategy, as the rewards are perceived to be fairly shared. Third, the domestic logistics enterprises are more sensitive to the information collaboration payoff distribution coefficient. When the information collaboration payoff distribution is unreasonable, domestic logistics enterprises converge to 0 more quickly. In other cases, the initial willingness of domestic logistics enterprises shows a downward trend until the willingness of cross-border e-commerce platforms and foreign logistics enterprises to collaborate on information reaches a certain threshold. At this point, when the loss from being excluded by alliance members for not cooperating exceeds the cost of information collaboration, the willingness of domestic logistics enterprises to engage in information collaboration begins to rise, ultimately converging to 1.

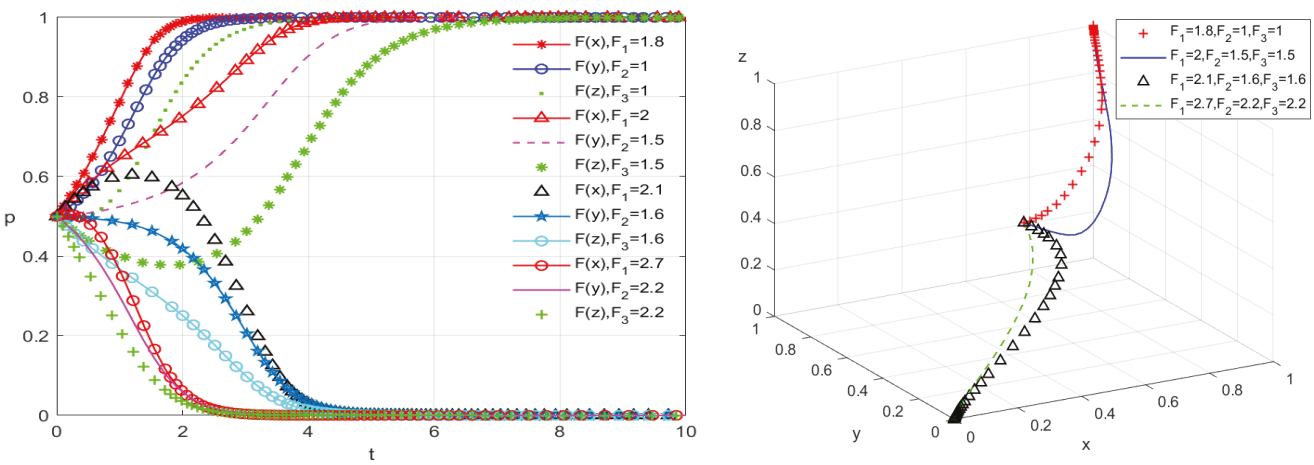
As shown in Figure 3b, when the information collaboration payoff distribution is more balanced, the convergence of the three players to the (1, 1, 1) strategy is faster, and their willingness to cooperate becomes stronger. This emphasizes that a more balanced payoff distribution fosters faster cooperation and enhances the likelihood that all parties will choose to collaborate. In contrast, when one player receives a disproportionately higher share of the payoff (such as when the cross-border e-commerce platform's share exceeds

0.6), the other members of the alliance tend to resist, making information collaboration more difficult to implement.

These results demonstrate that the structure of the revenue distribution significantly affects the adoption of the information collaboration strategy. When the risk levels and input strengths of the players are relatively similar, an average or moderate revenue distribution serves as an effective motivator for collaboration. However, when one player's share of the revenue is too high, it leads to resistance from the other players, which can hinder successful collaboration.

5.3. Effect of Information Collaboration Risk on Evolutionary Outcomes

Figure 4 shows the simulation of the effect of the information collaboration strategy among the three players when the cost of information collaboration risk takes different values.



(**a**) (**b**)

Figure 4. Effect of information collaboration risk on tripartite evolutionary outcomes. (**a**) Two-dimensional evolutionary paths. (**b**) Three-dimensional evolutionary paths.

As shown in Figure 4a, the critical change value of F_1 is between 2.0 and 2.1. When F_1 is below the critical value, x converges to 1; when F_1 exceeds the critical value, x converges to 0. The critical values of F_2 and F_3 range from 1.5 to 1.6. When F_2 and F_3 are below the critical value, y and z converge to 1. However, if F_3 does not reach a sufficiently small value, the willingness of domestic logistics enterprises to collaborate will decline until the collaboration willingness of cross-border e-commerce platforms and foreign logistics enterprises rises to a certain level. When F_2 and F_3 exceed the critical value, y and z converge to 0, with z converging to 0 faster than y. This indicates that domestic logistics enterprises are more sensitive to the risk of information synergism, as their cross-border logistics businesses are still in the early stages and cannot bear the associated risks. When F_i exceeds the critical value, x, y, and z all converge to 0, and the equilibrium point tends toward (0, 0, 0). At this point, an increase in F_i causes the three players to converge to 0 more quickly. When F_i is below the critical value, x, y, and z converge to 1, and the final equilibrium point tends toward (1, 1, 1). As F_i increases, x, y, and z converge to 1 more slowly. As shown in Figure 4b, as the risk of information collaboration increases, the tripartite evolutionary path evolves from (0, 0, 0) to (1, 1, 1) and eventually stabilizes.

5.4. Analysis of Results

This study constructs a tripartite evolutionary game model for information collaboration among cross-border e-commerce platforms, domestic logistics enterprises, and foreign logistics enterprises. The stability of their information collaboration strategy selection and

the factors influencing this collaboration are analyzed. The results and recommendations are presented based on the analysis and comparison, considering the following aspects:

(1) The study demonstrates that when the level of information collaboration standardization exceeds a certain threshold (e.g., 0.34), the benefits of information sharing become significantly more pronounced. This, in turn, encourages stronger collaboration among the involved parties. To leverage this finding effectively, cross-border e-commerce platforms should take a proactive role in accelerating the development of standardized information collaboration systems. Governments, both in China and internationally, can further support these efforts by investing in digital infrastructure and implementing policies that promote the development of emerging technologies such as blockchain and artificial intelligence (AI) to enhance data sharing. For instance, the European Union's Digital Single Market initiative aims to harmonize digital regulations and facilitate cross-border data flow, thereby improving e-commerce logistics efficiency. Similarly, governments in other regions could foster innovation by creating incentives—such as tax breaks for tech startups or public–private partnerships in logistics innovation—to encourage the adoption of new business models and services. Additionally, governments should prioritize unifying the information systems of key enterprises to ensure seamless logistics data exchange and electronic document flow. This could involve standardizing data formats and protocols to enable easier integration across borders. By leveraging the transparency of online platforms and integrating them with offline businesses, talent hubs, and resource innovation initiatives, governments can help build a robust and sustainable cross-border e-commerce information collaboration system. Such a system would foster long-term growth and enhance the competitiveness of global e-commerce.

(2) Compared to other influencing factors, alliance members exhibit the fastest convergence speed in information collaboration strategy selection under the influence of information collaboration risk. This indicates that information collaboration risk is the most significant factor in the evolution of cross-border logistics alliance members' information collaboration decision-making. Cross-border e-commerce platforms and logistics involve different countries and regions, necessitating compliance with the governmental requirements and regulations of various nations. As a result, stable and clear policies, such as bilateral trade agreements and the standardization of customs procedures, will help foster cooperation between cross-border enterprises. For instance, platforms can collaborate with regulatory bodies to ensure compliance with international data protection laws like GDPR in the EU or CCPA in California, thereby reducing legal risks and encouraging trust among alliance members. At the same time, alliance members should strengthen the prevention and control of information collaboration risks by developing risk management frameworks. For example, they could implement robust cybersecurity protocols, such as end-to-end encryption and regular audits, to safeguard data exchanges. Additionally, they should work towards harmonizing cultural and institutional differences by conducting cross-cultural training and establishing clear communication protocols to minimize misunderstandings. To reduce errors in information standards, enterprises could adopt internationally recognized standards, such as the GS1 global standards for barcoding and data synchronization, ensuring compatibility in information exchange and minimizing errors during integration. In summary, information collaboration risk proves to be the most influential factor in the evolution of collaboration strategies among cross-border logistics alliance members. To address this practical implication, stable and clear policies—such as bilateral trade agreements and the standardization of customs procedures—are essential for fostering collaboration. Platforms can also work with regulatory bodies to ensure compliance with international data protection laws (e.g., GDPR in the EU or CCPA in California) to mitigate legal risks and enhance trust. Additionally, strengthening risk man-

agement frameworks is crucial for preventing cybersecurity threats and ensuring smooth data exchange. Companies could adopt internationally recognized standards, such as GS1 for barcoding and data synchronization, which would minimize errors during information exchange and integration. Governments can support this process by providing guidelines and certifications for cross-border logistics collaborations.

(3) In terms of payoff distribution, the average payoff distribution model is more effective in promoting the selection of the information collaboration strategy. However, if the payoff distribution coefficient for the cross-border e-commerce platform exceeds 0.6, it will be resisted by other alliance members, making the implementation of information collaboration difficult. To avoid this imbalance, it is essential to assess the information collaboration risks and contributions of each alliance member. A clear and fair payoff distribution mechanism should be developed based on this assessment. For example, the risk and reward sharing model could involve calculating the input of each member based on factors such as data sharing volume, system integration costs, and compliance with international standards. Cross-border e-commerce platforms could share the risk by offering smaller platforms or logistics partners a proportional share of the overall benefit, thus preventing any one party from dominating the collaboration. In practice, companies could adopt a transparent model where each member's contribution to risk mitigation (e.g., implementing robust cybersecurity protocols or adopting standardized data formats) is assessed and rewarded. For instance, an alliance could introduce a "performance-based reward system", where logistics companies that invest in real-time tracking technologies or automated customs processing systems receive a higher share of the payoff, reflecting their contributions to improving the efficiency of information flow and reducing operational risks. Moreover, the payoff distribution model should be flexible enough to adapt to changes in the level of collaboration, such as new members joining the alliance or changes in the regulatory environment, to ensure it remains fair and aligned with each party's contribution. The study highlights that a moderate payoff distribution is most effective in fostering cooperation among alliance members. Conversely, a highly imbalanced distribution (e.g., where the e-commerce platform's share exceeds 0.6) can lead to resistance. To prevent such imbalances, a fair and transparent revenue-sharing model should be established, based on thorough risk and contribution assessments. For example, logistics companies that invest in advanced technologies (e.g., real-time tracking or automated customs processing) could receive a larger share of the payoff, reflecting their added value.

To ensure long-term sustainability, the payoff distribution model should also remain adaptable. For instance, adopting a performance-based reward system could incentivize better collaboration by recognizing each member's efforts in mitigating risks or improving operational efficiency. This approach not only promotes fairness but also strengthens the alliance's overall resilience and effectiveness.

(4) The results for the effect of the three factors on the evolutionary game show that convergence speed is the fastest under the influence of information collaboration risk; t = 2 quickly converges to 0 or 1, indicating that the three players are highly sensitive to the effect of information collaboration risk. Therefore, to enhance the willingness of cross-border logistics enterprises to collaborate, it is crucial to implement targeted supervision and conduct a thorough risk assessment. One specific action would be to establish a risk management framework that identifies key risks associated with information sharing, such as data breaches, cybersecurity threats, or non-compliance with international regulations (e.g., GDPR, CCPA). For example, the alliance could introduce regular cybersecurity audits and establish compliance checkpoints to ensure that all members adhere to agreed-upon data protection standards, thus reducing collaboration risks. Another practical action is to create a "collaboration risk scorecard", which evaluates the risk tolerance and contri-

butions of each member based on their compliance with information security protocols, data sharing practices, and adherence to international standards. The scorecard could assign risk scores to each member, with those who meet higher standards receiving greater benefits or a larger share of the rewards from the collaboration. Furthermore, governments or industry bodies could play a supporting role by providing guidelines and certifications for cross-border logistics collaboration, ensuring that all parties are aligned on minimum standards and practices. For instance, a government could establish a certification program that rewards companies for adopting standardized information exchange practices, such as using blockchain for tracking or adopting common data formats like GS1, which can reduce errors and risks in information exchanges. The analysis highlights that the convergence speed of collaboration strategies is fastest when influenced by information collaboration risk. This underscores the importance of creating targeted risk management frameworks. Regular cybersecurity audits, compliance checkpoints, and the development of a collaboration risk scorecard to assess risk tolerance and security protocol compliance can incentivize the adoption of best practices. Additionally, governments or industry bodies can support these efforts by offering certifications for companies that meet minimum information exchange standards, such as using blockchain for tracking.

(5) In addition, this study found that whether the participants choose information collaboration is generally not affected by their initial income; however, their initial income does influence their total income. When a participant's daily operating income is stable, the information collaboration standard, risk, and payoff distribution have a greater impact on the participant's strategy choice. Therefore, cross-border e-commerce platforms and logistics enterprises should consider the stability of their partners' financial positions when forming alliances, ensuring that less financially stable partners are not excluded or left behind in the collaborative process. By fostering inclusive partnerships and managing risks effectively, they can maximize the overall benefits of information collaboration.

6. Conclusions

The use of the evolutionary game theory in this study offers a novel perspective on information collaboration within cross-border logistics alliances. The findings significantly advance our theoretical understanding of how the interplay between information collaboration risks and payoff distribution influences decision-making within these alliances. Specifically, this research highlights the importance of strategic cooperation among e-commerce platforms, domestic logistics companies, and international logistics firms, and how their interaction shapes the overall performance of cross-border logistics networks.

However, cross-border logistics alliances involve a wide range of stakeholders with differing interests and operate in complex market environments. These complexities may limit the generalizability of our findings. To expand on these insights and enhance the robustness of the conclusions, future research should explore several key areas. Firstly, incorporating additional factors such as consumer preferences, market size, and industry-specific characteristics could provide a more granular understanding of the decision-making behaviors of alliance members. This would allow for the development of more tailored payoff distribution models that better align with the diverse interests of both logistics providers and end consumers. Furthermore, applying this model to real-world case studies across various sectors and regions would offer valuable empirical evidence, testing the model's applicability in different market conditions and providing practical insights into the implementation of information collaboration strategies. Secondly, future research should consider the impact of external factors such as government policies, technological advancements, and global trade regulations on the dynamics of cross-border logistics alliances. Incorporating these variables would offer a more comprehensive view of the mechanisms

that underpin sustainable and long-term cooperation within these alliances. Moreover, exploring the role of digital platforms and emerging technologies—such as blockchain or artificial intelligence—could present innovative solutions to enhance information sharing, risk mitigation, and operational efficiency in cross-border logistics.

In conclusion, while this study provides valuable theoretical insights into the evolution of information collaboration in cross-border logistics alliances, there is considerable potential for future research to further refine and expand the findings. By exploring additional variables and real-world applications, future studies can contribute to the development of more effective, stable, and sustainable models for cross-border logistics cooperation.

Author Contributions: Conceptualization, X.M. and Z.L.; methodology, X.M. and Y.Y.; software, X.M.; validation, X.M., Z.L., and A.X.; formal analysis, X.M.; investigation, X.M. and A.X.; resources, X.M.; data curation, X.M.; writing—original draft preparation, X.M.; writing—review and editing, X.M. and Y.Y.; visualization, X.M.; supervision, Z.L.; project administration, Z.L.; funding acquisition, A.X. All authors have read and agreed to the published version of the manuscript.

Funding: The China National Social Science Foundation (21BGL148), Preliminary Research Projects for Fujian Province's Tenth Five-Year Plan.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors thank the anonymous referees for their invaluable comments on an earlier version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, Y.; Wang, Y.; Lee, S.H. The effect of cross-border e-commerce on China's international trade: An empirical study based on transaction cost analysis. *Sustainability* **2016**, *9*, 2028. [CrossRef]
2. Cucchiella, F.; Gastaldi, M. Risk management in supply chain: A real option approach. *J. Manuf. Technol. Manag.* **2006**, *17*, 700–720. [CrossRef]
3. Shen, L.; Lin, F.; Wang, Y.Y. Decision and Coordination of an Ecommerce Supply Chain Considering Returns and Net-work Externalities. *J. Syst. Sci. Syst. Eng.* **2024**, *33*, 552–575. [CrossRef]
4. Togar, M.S.; ALAN, C.W.; Sridaran, R. The knowledge of coordination for supply chain integration. *Bus. Process Manag. J.* **2002**, *8*, 289–308. [CrossRef]
5. Togar, M.S.; ALAN, C.W.; Sridaran, R. Applying the theory of constraints to supply chain collaboration. *Int. J. Supply Chain Manag.* **2004**, *9*, 57–70. [CrossRef]
6. Togar, M.S.; ALAN, C.W.; Sridaran, R. An integrative framework for supply chain collaboration. *Int. J. Logist. Manag.* **2005**, *16*, 257–274. [CrossRef]
7. Tenorio, M.; Pascucci, S.; Verkerk, R.; Dekker, M.; van Boekel, T.A. What Does It Take to Go Global? The Role of Quality Alignment and Complexity in Designing International Food Supply Chains. *Supply Chain Manag.* **2021**, *26*, 467–480. [CrossRef]
8. Ye, W. E-Commerce Logistics and Supply Chain Network Optimization for Cross-Border. *J. Grid Comput.* **2024**, *22*, 22. [CrossRef]
9. Zhang, X. Prediction of Purchase Volume of Cross-Border e-Commerce Platform Based on BP Neural Network. *Comput. Intell. Neurosci.* **2022**, *2022*, 3821642. [CrossRef] [PubMed]
10. Wu, P.J.; Lin, K.C. Unstructured big data analytics for retrieving e-commerce logistics knowledge. *Telemat. Inform.* **2018**, *35*, 237–244. [CrossRef]
11. Yang, X.C.; Jiang, H.Y.; Chen, W.X. Evolutionary Game Analysis of Cross-Border E-Commerce Logistics Alliance Subject Considering Supply Chain Disruption Risk. *Sustainability* **2023**, *15*, 16350. [CrossRef]
12. Zhang, X.M.; Zha, X.Y.; Zhang, H.Y.; Dan, B. Information Sharing in A Cross-Border E-Commerce Supply Chain Under Tax Uncertainty. *Int. J. Electron. Commer.* **2022**, *26*, 123–146. [CrossRef]
13. Gurbanova, N.; Wang, Z. The Chinese Experience of Developing Cross-Border E-Commerce Under The "Belt and Road" Framework: Factors, Challenges, and Solutions. *J. Glob. Inf. Manag.* **2023**, *31*, 21. [CrossRef]

14. Zhang, J.; Zhang, Q.; Bai, Z.Y. Coordinating A Supply Chain for Deteriorating Items with A Revenue Sharing and Cooperative Investment Contract. *OMEGA-Int. J. Manag. Sci.* **2015**, *56*, 37–49. [CrossRef]
15. Du, Z.P.; Fu, S.S.; Mu, D. Multiparty Behavior Game Research of Cross-border E-commerce Logistics Alliance Based on 4PL. *Chin. J. Manag. Sci.* **2020**, *28*, 104–113. [CrossRef]
16. Wang, D.L.; Li, W.W. Optimization algorithm and simulation of supply chain coordination based on cross-border E-commerce network platform. *Wirel. Mob. Comput.* **2021**, *23*, 920–934. [CrossRef]
17. Barut, M.; Faisst, W.; Kanet, J.J. Measuring supply chain coupling: An information system perspective. *Eur. J. Purch. Supply Manag.* **2002**, *8*, 161–171. [CrossRef]
18. Cui, Y.; Huang, M.; Yang, S. Fourth Party Logistics Routing Problem Model with Fuzzy Duration Time and Cost Discount. *Knowl.-Based Syst.* **2013**, *50*, 14–24. [CrossRef]
19. Dai, P.; Xu, J.; Li, W. Research on Profit Distribution of Logistics Alliance Considering Communication Structure and Task Completion Quality. *Processes* **2022**, *10*, 1139. [CrossRef]
20. Chen, S.H.; Xiao, H.; Huang, W.Z.; He, W. Cooperation of Cross-border E-commerce: A Reputation and Trust Perspective. *J. Glob. Inf. Technol. Manag.* **2022**, *25*, 7–25. [CrossRef]
21. Zhou, L.; Wang, J.J.; Li, F.; Xu, Y.; Zhao, J.; Su, J. Risk Aversion of B2C Cross-Border E-Commerce Supply Chain. *Sustainability* **2022**, *14*, 8088. [CrossRef]
22. Liu, L.; Jiang, X.Y.; Wu, J. Research on The Construction of China-Europe Cross-border E-commerce Ecosphere Based on Blockchain Technology. *J. Financ. Res.* **2019**, *3*, 11. [CrossRef]
23. Rashid, A.; Rasheed, R. Logistics Service Quality and Product Satisfaction in E—Commerce. *SAGE Open* **2024**, *14*, 21582440231224250. [CrossRef]
24. Wang, G.F.; Zhang, Z.H.; Li, S.; Shin, C. Research on The Influencing Factors of Sustainable Supply Chain Development of Agri-Food Products Based on Cross-Border Live-Streaming E-Commerce in China. *Foods* **2023**, *12*, 3323. [CrossRef] [PubMed]
25. Chen, Q.P. Analysis and Prediction of Cross-Border E-Commerce Scale of China Based on The Machine Learning Model. *Comput. Intell. Neurosci.* **2022**, *23*, 7906135. [CrossRef] [PubMed]
26. Mensah, I.K.; Zeng, G.; Luo, C. The effect of national culture dimensions on the adoption of cross-border E-commerce. *Int. J. Inf. Syst. Serv. Sect.* **2020**, *12*, 91–112. [CrossRef]
27. Min, H. Blockchain technology for enhancing supply chain resilience. *Bus. Horiz.* **2019**, *62*, 35–45. [CrossRef]
28. Liu, S. Multimodal Transportation Route Optimization of Cold Chain Container in Time-Varying Network Considering Carbon Emissions. *Sustainability* **2023**, *15*, 4435. [CrossRef]
29. Chen, T.G.; Qiu, Y.W.; Wang, B.; Yang, J. Analysis of Effects on The Dual Circulation Promotion Policy for Cross-Border E-Commerce B2B Export Trade Based on System Dynamics During COVID-19. *Systems* **2022**, *10*, 13. [CrossRef]
30. Qiu, Y.W.; Chen, T.G.; Cai, J.; Yang, J.J. The Impact of Government Behavior on the Development of Cross-Border E-Commerce B2B Export Trading Enterprises Based on Evolutionary Game in the Context of "Dual-Cycle" Policy. *J. Theor. Appl. Electron. Commer. Res.* **2022**, *17*, 1741–1768. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Big Data-Driven Carbon Trading and Industrial Firm Value Based on DEA and DID

Zhen Peng [1], Yunxiao Zhang [2] and Tongtong Sun [2,*]

[1] School of Economics and Management, China University of Geosciences Beijing, Beijing 100083, China
[2] School of Economics and Management, Beijing Institute of Petrochemical Technology, Beijing 102617, China
* Correspondence: suntongtong@bipt.edu.cn

Abstract: Carbon trading has emerged as a critical environmental and economic mechanism for promoting energy conservation and emission reduction among firms in China. Leveraging big data from listed industrial firms participating in carbon trading, this study employs the super-efficiency SBM model and the common frontier model to evaluate firm-level carbon performance. Using carbon performance as a mediating variable, the study investigates the impact of carbon trading on firm value, considering the moderating effects of internal and external governance mechanisms. The findings reveal the following: (1) Carbon trading enhances firm value by improving carbon performance. (2) Internal governance mechanisms strengthen the positive effect of carbon trading on firm value, while government intervention weakens this effect. (3) The value-enhancing effect of carbon trading is more pronounced for firms in China's central and western regions. (4) Among industrial firms, carbon trading has the strongest impact on the value of manufacturing firms. These results provide valuable insights for policymakers and firms aiming to align environmental and economic objectives through carbon-trading mechanisms.

Keywords: big data; carbon trading; industrial firm value; firm carbon performance; internal and external governance of firms

1. Introduction

In the quest for sustainable development, the intersection of environmental policy and industrial economics has given rise to innovative mechanisms aimed at mitigating climate change. Among these, carbon trading has emerged as a pivotal strategy to internalize the externalities of greenhouse gas emissions [1]. Globally, 28 carbon-trading markets currently cover approximately 17% of global carbon emissions, totaling 9 billion tons. In China, the advent of carbon-trading policy represents a significant step in aligning economic activities with environmental stewardship. The carbon-trading policy was implemented in Shenzhen, Shanghai, Beijing, Guangdong, and Tianjin in 2013 [2]; in Hubei and Chongqing in 2014; and in Fujian in 2016 [3]. China's carbon trading involves dozens of industries such as power, heat, cement, petrochemicals, manufacturing, public buildings, etc. Presently, over two thousand key emitting units are included in the national carbon-trading market, covering approximately 4.5 billion tons of carbon emissions. Industrial firms participate in carbon trading not only out of environmental responsibility but also due to the prospective economic benefits and growth opportunities it offers, as well as its potential impact on their market value [4].

Despite the growing body of research on carbon trading and its impact on firm value [5,6], the effectiveness of carbon trading in enhancing firm value remains a subject of considerable debate [7]. This debate is further complicated by the challenge of

quantifying policy impacts amidst a myriad of influencing factors. Existing studies often rely on limited data, resulting in generalized findings that fail to provide refined insights. Consequently, firms' ability to effectively address environmental pressures and mitigate risks is constrained.

The advent of big data has introduced a new paradigm in policy analysis, offering unprecedented insights into the microeconomic consequences of carbon-trading policy. Against this backdrop, this study aims to address the following research questions:

(1) How does carbon trading influence firm value through the improvement of carbon performance?
(2) What role do internal and external governance mechanisms play in moderating the relationship between carbon trading and firm value?
(3) Are there regional and industrial variations in the impact of carbon trading on firm value?

To answer these questions, this study constructs carbon performance indicators for industrial firms using the super-efficiency SBM and common frontier models. By examining the impact of carbon trading on firm value from multiple perspectives, this research provides stakeholders with more comprehensive evaluation metrics for corporate behavior. Furthermore, it offers practical insights for firms to effectively respond to environmental pressures, manage risks, and achieve sustainable development. For policymakers, the findings can inform the design of targeted incentive measures to enhance market activity and liquidity.

This study contributes to the existing literature in three key ways. First, it expands the understanding of the economic value of carbon trading at the firm level by leveraging big data. Second, it provides a nuanced analysis of the moderating effects of internal and external governance mechanisms. Third, it highlights regional and industrial variations in the impact of carbon trading, offering actionable insights for both firms and policymakers. By bridging the gap between academic research and corporate practice, this study aims to advance the discourse on sustainable development and the role of carbon trading in achieving environmental and economic objectives.

The remainder of this study is structured as follows. In Section 2, we briefly review carbon trading, carbon performance and firm value. Afterward, the detailed model of carbon performance, carbon-trading policy, and firm value is covered in Section 3. Experimental results are presented in Section 4 while conclusions are offered in Section 5.

2. Literature Review

After the implementation of carbon trading, scholars first focused on its impact on emission reduction [4,8–10]. As research continues to delve deeper, researchers have gradually shifted their focus to the dual value of carbon trading for both the economy and emission reduction. Consequently, studies on the impact of carbon trading on carbon performance have continued to be enriched. Carbon performance [8] refers to the evaluation results centered around a series of actions and outcomes that enterprises undertake in response to policy directives for green, low-carbon development, and the recycling of secondary raw materials, which involve controlling carbon-based energy consumption and reducing carbon emissions. Existing research on firm carbon performance typically links firm carbon emissions to financial performance to better measure the dual benefits of a firm's environmental and economic performance [11]. Good carbon performance has a positive effect on the market value of enterprises because investors tend to reward companies with low carbon emission levels [8].

From a regional perspective, Shi et al. [12] used the difference-in-differences (DID) method to assess the impact of carbon emission trading policy on regional carbon emission

performance and found that the reduction in carbon intensity depends on the intermediary effects produced by three factors: technological innovation incentives, industrial structure optimization, and energy substitution. Yu et al. [13] adopt the super-efficiency SBM model to evaluate provincial and municipal carbon performance and found that the carbon-trading mechanism effectively improved the carbon performance of pilot provinces and cities based on recursive models, and further explored the specific paths. Zhang et al. [14] studied the impact of the carbon-trading pilot policy on the Green Total Factor Productivity (GTFP) of listed manufacturing enterprises in China, and the results showed that carbon trading could effectively promote the GTFP of enterprises, among which Environmental, Social, and Governance (ESG) performance played a certain intermediary role. He et al. [15] examined the impact of carbon trading on the Total Factor Pollution Control Efficiency (TFPCE) of firms in the carbon-trading system, and the results showed that the carbon-trading policy significantly improved the TFPCE of carbon emission firms in the pilot areas, and also produced a green development effect, among which strategic innovation, substantive innovation, and institutional quality, respectively, played a positive role in improving pollution control performance. Yu et al. [16] found that the perfection degree of carbon market policy has a threshold effect on firm environmental performance, and there is an optimal interval for the perfection of carbon market policy.

Numerous scholars have conducted research on firm value. Commonly used quantitative methods in existing research are asset returns [17], or Tobin's Q value [18], etc. Some scholars have conducted research on the direct impact of carbon trading on corporate value. For instance, Han [19] indicates that carbon pricing significantly affects the value of both traditional and new energy companies, but the overall impact is relatively small, with a greater effect on short-term value than on long-term value. Lin [20] demonstrates that the carbon-trading system significantly enhances the firm value of listed firms in key industries, with a more pronounced effect on state-owned enterprises. Niu et al. [21] suggest that the carbon-trading system has a more significant role in promoting the market value of non-state-owned and larger enterprises. Yuan [22] shows that carbon trading strengthens a firm's market power through low-carbon innovation and market consolidation effects. In addition, other studies have shown that carbon trading can affect a firm's profitability [23], investment decisions [24], and the ability to transition into diversification [25], among other aspects, which in turn impact firm operational earnings and market value, resulting in changes in firm value.

The level of internal and external governance of firms has a significant impact on the effect of environmental regulations on firm value. Internal governance elements usually include shareholders, boards of directors, management, supervisory boards, etc. [26]. Xu et al. [27] have delved into the role of director and officer liability insurance in firm performance and value creation, revealed that by insulating directors and senior managers from legal liabilities arising from improper decisions, director and officer liability insurance helps to foster a more robust decision-making environment, which in turn positively enhances a firm's market value. However, due to the special attributes of state ownership and the high concentration of equity, the positive impact of director and officer liability insurance on firm value may be diminished. Once equity concentration exceeds a certain threshold and reaches the level of absolute control, it may lead to an increase in insider control and a failure of supervisory mechanisms, ultimately bringing negative consequences to the firm's market value.

At the same time, government intervention, as an important external governance factor, not only affects the enthusiasm of firms to comply with environmental regulations but also influences the sustainability of internal decision-making and the effectiveness of decision implementation [28]. Government intervention in firm development does not

necessarily have a positive effect; some studies have shown that excessive government intervention can deteriorate firm financial performance [29]. Although high levels of government intervention provide more loans and investments to firms, these firms may prioritize alignment with government objectives rather than maximizing shareholder wealth. They are more likely to implement government stimulus plans, even if investment projects are not profitable. Therefore, these firms often perform poorly after receiving investment, and investment efficiency decreases.

Existing academic research reveals that while studies on carbon performance are continually deepening, particularly in exploring the macro carbon emission performance at the regional level and the impact of carbon-trading systems on overall performance, there is a noticeable deficiency in the construction of carbon performance indicator systems at the micro firm level and detailed studies on the specific impacts of carbon-trading activities.

Furthermore, current research often includes rather general and not well-refined factors when discussing the differentiated impact of carbon trading on firm value, which greatly limits the precision of research conclusions in corresponding to and guiding the practical application of enterprises within the context of China's conditions. In light of this, there is an urgent need for further in-depth and detailed research on the specific pathways and extent of the influence of carbon trading on firm value through its complex internal mechanisms, in order to propose more targeted and applicable strategic recommendations.

With the advancement of big data, the analysis and modeling of many micro-level entities have become possible [30,31]. This paper offers unprecedented insights into the microeconomic consequences of carbon-trading policy. Utilizing a combination of Data Envelopment Analysis (DEA) and difference-in-differences (DID), it seeks to provide a comprehensive assessment of the influence of carbon-trading policy on firm value, while accounting for the mediating effects of corporate governance and government intervention.

3. Materials and Methods

3.1. Modeling Carbon Performance of Industrial Firm Based on DEA

As an extension of DEA (Data Envelopment Analysis), the super-efficiency SBM (Surface-Based Measure) model has a unique advantage. It can independently and finely discriminate and rank decision-making units whose efficiency scores have reached the optimal level (i.e., the efficiency value is 1) regardless of the reference group. The model can be configured into three different operational modes, among which, the non-oriented mode can comprehensively consider both input and output dimensions to measure the level of efficiency. Therefore, this paper uses the non-oriented mode to measure the carbon performance of industrial enterprises.

3.1.1. Feature Selection of Carbon Performance Modeling

Carbon performance indicators for industrial firms are shown in Table 1.

Table 1. Industrial carbon performance indicators.

Project	First-level indicator	Secondary indicators
Inputs	Capital investment	Net value of fixed assets
	Labor input	Total number of employees
	Energy inputs	Total energy consumption
Outputs	Expected outputs	Main business income
		Total profits and taxes
	Non-expected output	Carbon dioxide emissions

(1) Input features

Based on Qin et al. [32], the net value of fixed assets represents the capital investment of firm. The total number of employees, following the literature [33], serves as a measure for the firm's labor input. To accurately measure the difference in energy consumption among different listed firms, various types of energy consumption are converted into a unified standard coal (unit: tons of standard coal) using the conversion coefficient [34].

$$\textit{Unified Standard Coal} = \textit{Water Consumption} \times 0.0002429 + \textit{Electricity Consumption} \times 1.229 + \textit{Raw Coal Usage} \times 0.7143 + \textit{Natural Gas Usage} \times 13.3 + \textit{Gasoline Usage} \times 1.4714 + \textit{Diesel Usage} \times 1.4571 + \textit{Central Heating} \times 0.03412$$

(2) Output features

Following the research [35], the main expected output of industrial firms is measured using the firm's main business income and total profits and taxes. The non-expected output, namely carbon dioxide emissions of industrial firms, is calculated as follows [36]:

$$CE = \sum \textit{Firm Main Energy Consumption} \times \textit{Carbon Dioxide Emission Coefficient}$$

3.1.2. Carbon Performance Model Building

The mathematical expression for the super-efficiency SBM model is as follows:

$$\rho = min \frac{1 + \frac{1}{M}\Sigma_{i=1}^{M}\left(s_i^x / x_{i0}^t\right)}{1 - \frac{1}{R+J}\left[\Sigma_{k=1}^{R}\left(s_k^y / y_{k0}^t\right) + \Sigma_{l=1}^{J}\left(s_l^b / b_{l0}^t\right)\right]},$$

subject to

$$\begin{cases} x_{i0}^t \geq \sum_{t=1}^{T}\sum_{j=1,\neq 0}^{n}\lambda_j^t x_{ij}^t + s_i^x, i = 1,2,...M, \\ y_{k0}^t \leq \sum_{t=1}^{T}\sum_{j=1,\neq 0}^{n}\lambda_j^t y_{kj}^t - s_k^y, k = 1,2,...R, \\ b_{l0}^t \geq \sum_{t=1}^{T}\sum_{j=1,\neq 0}^{n}\lambda_j^t b_{l0}^t + s_j^b, l = 1,2,...J, \\ 1 - \frac{1}{R+J}\left[\sum_{k=1}^{R}\left(s_k^y / y_{k0}^t\right) + \sum_{l=1}^{J}\left(s_l^z / b_{l0}^t\right)\right] > 0, \\ s_i^x \geq 0, s_k^y \geq 0, s_l^b \geq 0, , \forall i, , k, l, \\ \sum_{t=1}^{T}\sum_{j=1,\neq 0}^{n}\lambda_j^t = 1; \lambda_j^t \geq 0, \end{cases} \tag{1}$$

where s_i^X, s_k^y, s_l^b represent input, expected output, and the degrees of relaxation for non-expected output, respectively. The weight coefficient vector λ is used to characterize the importance of each evaluated unit. The subscript "0" in the model specifically refers to the particular individual being evaluated. The result of the objective function calculation can be interpreted as the carbon performance of the industrial enterprise being evaluated.

Due to the dual impact of economic and policy factors, the phenomenon of industrial agglomeration in China is becoming increasingly evident. The geographical distribution of industrial enterprises shows a clear trend of clustering, and according to the regional delineation laws (Several Opinions of the Central Committee of the Communist Party of China and Opinions of The State Council on the Implementation of Several Policies and Measures for the Development of the Western Region), the overall region of China can be divided into four clusters: east, central, west, and northeast [37]. By analyzing the gap between the frontier surfaces constituted by industrial enterprises in different regions and the common frontier, we can better analyze the difference in technological level between the regions where industrial enterprises are located and the optimal technological level across the entire region. The efficiency value of the common frontier is equivalent to the efficiency

value of the overall super-efficiency SBM model without grouping. In the common frontier model, the TGR (Technology Gap Ratio) can be expressed by the following formula:

$$\text{TGR} = \frac{Group\ Frontier\ Efficiency\ Value}{Common\ Frontier\ Efficiency\ Value} \tag{2}$$

Since the common frontier efficiency value is less than or equal to the group frontier efficiency value, the TGR should be less than or equal to 1. The smaller the TGR, the greater the technological gap between the group's group technology and the overall technology; if the TGR is 1, it indicates that the group's group technology coincides with the overall technology, suggesting that the group is at the forefront of technological innovation.

3.2. Modeling the Impact of Carbon Trading on Industrial Firm Value Based on DID

It can be observed that carbon trading in China has been established in various provinces over time. Since the impact of the carbon-trading intervention occurs at multiple different points and continues to exist from the time of implementation, the traditional difference-in-differences (DID) model is not suitable. The multi-time-point DID model can address the issue of different policy implementation times across regions.

3.2.1. Feature Selection of Impact Modeling

(1) Launch of the carbon emissions trading policy

Since China planned to launch a pilot carbon emission trading market in 2011, pilot areas have successively established carbon-trading markets. In 2013, Shenzhen, Shanghai, Beijing, Guangdong, and Tianjin officially started carbon-trading markets. In 2014, Hubei and Chongqing carbon-trading markets were initiated. In 2016, the Fujian carbon-trading market was launched. The years 2013, 2014, and 2016 are taken as policy shock years, respectively, and included as dummy variables in the model.

(2) Industrial firm value

Tobin's Q indicator, based on the market value of firm capital and replacement capital, including both liabilities and equity, is a comprehensive indicator capable of measuring firm value and future potential widely used in various studies [38]. This paper employs Tobin's Q value as the measure of firm value, providing insights into the impact of investment decisions.

(3) Internal firm governance

Drawing from Chen [39], the proxy variable for the internal firm governance level is ownership concentration, expressed as the proportion of the top ten shareholders' shares to the firm's total shares.

(4) External firm governance

Following the literature [40], the external governance level is determined by the degree of government intervention, i.e., the proportion of general budget expenditures in GDP in the province where the firm is located.

3.2.2. Impact Model Building

To investigate the possible mediating effect of firm carbon performance between carbon trading and firm value, the following mediation model is constructed and tested using Sobel. The mediation model constructed by stepwise regression method is as follows.

$$\text{CP}_{i,t} = \theta_0 + \theta_1 \times \text{CET}_{i,t} + \Sigma\text{Control}_{i,t} + \varepsilon_{i,t} \tag{3}$$

$$TQ_{i,t} = \eta_0 + \eta_1 \times \text{CET}_{i,t} + \eta_2 \times \text{CP}_{i,t} + \Sigma\text{Control}_{i,t} + \varepsilon_{i,t} \tag{4}$$

In addition, we also tested the moderating effect of internal and external firm governance elements in the linear relationship between carbon trading and firm value by including carbon-trading policy implementation as an explanatory variable in the model, respectively, and constructing Equation (5).

$$TQ_{i,t} = g_0 + g_1 \times CET_{i,t} + g_2 \times CET_{i,t} \times Own_{i,t} + g_3 \times Gov_{i,t} + g_4 \times Own_{i,t} \times Gov_{i,t} + g_5 \times CET_{i,t} \times Gov_{i,t} + g_6 \times CET_{i,t} \times Own_{i,t} \times Gov_{i,t} + g_7 \times Own_{i,t} + \Sigma Control_{i,t} + \varphi_i + \psi_t + \varepsilon_{i,t} \quad (5)$$

The coefficient in (5) g_2 reflects the degree of moderation of the internal governance element in the path of carbon-trading policy implementation on firm value, the coefficient g_5 reflects the degree of regulation of external governance elements, and the coefficient g_6 reflects the combined synergistic moderating effect of the internal and external governance elements of the firm.

3.3. Data Descriptive Analysis

In accordance with the research [39], this paper selects firm development scale (Size), firm sustainable development capability (Sustain), firm operation and profitability (Operation), and innovation level (Innovation) as control variables in the empirical model. To better reflect the macro background of firm value changes, this paper also introduces macro variables such as energy industry classification (Traditional Energy Industry) and industrialization level (Industrialization Level) as control variables. Relevant financial data are sourced from the CSMAR database, while other data come from annual reports and social responsibility reports disclosed by listed firms. Industrialization level is extracted from the China Industrial Statistical Yearbook. Data specifications are presented in Table 2.

Table 2. Feature descriptions.

Features	Symbol	Indicator Calculation
Firm value	TQ	Tobin's Q index = market value at the end of the year/total assets at the end of the year
Carbon-trading policy implementation	CET	Dummy variable for whether the firm's location implemented carbon-trading policy in the year: 1 for yes, 0 for no
Carbon performance	CP	Common frontier efficiency value derived from super-efficiency SBM model
Internal firm governance level	Own	Ownership concentration ratio = the proportion of shares held by the top ten shareholders to the firm's total shares
External firm governance level	Gov	The proportion of general budget expenditure to GDP in the province where the firm is located
Firm development scale	Fds	Ln (total assets at the end of the year)
Firm sustainable development capabilities	Fsc	Capital accumulation rate = growth in owner's equity/owner's equity at the beginning of the year × 100%
Business operations and profitability	Bop	Return on assets = profit before interest and tax/average total assets
Firm innovation level	Fil	Firm innovation level = ln (number of green invention patent authorizations + number of green utility model patent authorizations + 1)
Energy industry level	Eil	When the firm's industry is coal mining and washing oil (B06), natural gas mining (B07), petroleum processing, coking and nuclear fuel processing in manufacturing (C25), electricity and heat, production and supply, and gas production and supply (D44 and D45), this value is 1; otherwise, it is 0
Industrialization level	Il	The industrial added value of the firm's location as a proportion of the gross product of the province or region where the firm is located

In this research, we analyze a total of 10,061 samples from 1062 listed industrial firms over the period from 2010 to 2019. Descriptive statistics for each indicator, as shown in Table 3, are calculated. They include mean, median, standard deviation, minimum, and maximum values, which summarize the characteristics of the sample firms.

Table 3. Descriptive statistical results.

Features	Number of Samples	Mean	Median	Standard Deviation	Minimum Value	Maximum Value
CET	10,061	0.240	0	0.427	0	1
TQ	10,059	2.016	1.619	1.280	0.699	19.82
CP	10,061	0.175	0.110	0.233	0.000	4.367
Own	10,061	56.63	56.67	15.00	8.970	100.0
Gov	10,061	0.193	0.173	0.0740	0.113	0.758
Fds	10,061	22.26	22.085	1.251	18.760	28.636
Fsc	10,061	0.245	0.0640	1.266	−1.415	98.08
Bop	10,061	0.055	0.0480	0.0620	−1.029	0.767
Fil	10,061	2.175	0	12.06	0	512
Eil	10,061	0.988	1	0.109	0	1
Il	10,061	0.383	0.401	0.0810	0.111	0.530

4. Results

4.1. Carbon Performance Measurement and Analysis of Industrial Firms

In this study, 1062 listed industrial enterprises are selected for carbon performance measurement. The analysis covered 10 time windows from 2010 to 2019, with the assumption of variable returns to scale. Based on Chinese policies, the firms are divided into four regional clusters: east, central, west, and northeast. Group frontier surfaces are constructed for each cluster to calculate the carbon efficiency values of industrial firms over different periods. Additionally, a common frontier surface encompassing all clusters is constructed to derive the carbon efficiency values under the common frontier. The TGR (Technology Gap Ratio) is then calculated by comparing the efficiency values of each cluster with the common frontier.

As shown in Table 4, the carbon performance measurements reveal that the firms with the highest carbon efficiency values are concentrated in the eastern region. The TGR values of firms in the eastern region are close to 1, indicating a balanced technology level and strong technological innovation capabilities. This technological balance facilitates the formation of industrial agglomeration effects [41], enabling firms to better address challenges related to carbon emissions and resource utilization. In contrast, the western region exhibits higher TGR values, suggesting a need for enhanced environmental protection and carbon emission reduction measures. The central and northeastern regions show lower carbon efficiency, which can be attributed to their traditional industrial structures and relatively lower levels of economic development.

In summary, the eastern region excels in carbon efficiency and technological innovation, requiring continued support for sustainable development. The western region shows promise but needs environmental improvement. The central and northeastern regions need increased investment in innovation and environmental protection.

Table 4. Top carbon performance results of industrial firms in China.

Ranking	Company Name	Region	Common Frontier Carbon Efficiency	Group Frontier Carbon Efficiency	TGR
1	** Iron&Steel Co., Ltd. in Hangzhou of China	eastern	4.3669	4.3669	1
2	** Heng Group in Wuzhou of China	western	3.1494	3.5231	0.8939
3	Shang ** Motor Co., Ltd. in Shanghai of China	eastern	2.4923	2.5173	0.9901
4	Hong ** Mining Co., Ltd. in Zibo of China	eastern	2.4175	2.4175	1
5	Chinese ** Petroleum Co. in Beijing of China	eastern	2.2522	2.2522	1
6	Na ** Da Co. in Zhuhai of China	eastern	2.2212	2.3456	0.9469
7	** Wen Technology in Kunming of China	western	2.1587	2.9431	0.7335
8	Shanghai ** Co., Ltd. in Shanghai of China	eastern	1.9720	1.9720	1
9	** Tai Power Co., Ltd. in Shanghai of China	eastern	1.8551	1.8551	1
10	Lao ** Xiang Co., Ltd. in Shanghai of China	eastern	1.7370	1.7370	1

4.2. Impact Analysis of Carbon Performance in Carbon Trading and Industrial Firm Value

The results in Table 5 show that the implementation of carbon-trading policy in China improves firm value by improving firm carbon performance. There are two reasons.

Table 5. Mediating effect of carbon trading on industrial firm value.

	CP	TQ
CET	0.042 ***	0.340 ***
	(0.005)	(0.029)
CP		0.577 ***
		(0.053)
Own	0.002 ***	−0.005 ***
	(0.000)	(0.001)
Fds	0.037 ***	−0.426 ***
	(0.002)	(0.010)
Fsc	0.010 ***	−0.040 ***
	(0.002)	(0.009)
Bop	0.918 ***	4.320 ***
	(0.035)	(0.193)
Fil	0.018 ***	0.029 **
	(0.003)	(0.015)
Eil	0.079 ***	0.080
	(0.020)	(0.105)
Il	−0.136 ***	−0.969 ***
	(0.029)	(0.156)
_cons	−0.829 ***	11.638 ***
	(0.050)	(0.272)
N	10,059	10,059
R^2	0.163	0.202
F	244.69	282.42

: 5% significance level ($p < 0.05$); *: 1% significance level ($p < 0.01$).

On one hand, as the total emissions are set, determining the scarcity of emission quotas, carbon emissions become an external negative factor in industrial operations. With the implementation of carbon trading, firms subject to carbon emission constraints will incur certain costs, leading them to emphasize carbon reduction. For industrial firms with high

energy consumption and severe pollutant emissions, constrained by the limit of carbon emission quotas, they will actively seek low-carbon energy to replace their existing energy sources, thereby improving their carbon performance [42]. For firms with lower pollution emissions, they can sell excess carbon emission quotas on the secondary carbon emission quota market at a reasonable price and use the proceeds to improve processes and reduce funds and labor [4], thereby improving their carbon performance.

On the other hand, information on industrial firms' carbon performance is considered internal information that is difficult for external investors to obtain. When a firm's carbon performance level is good, it is more inclined to disclose internal carbon performance information in social responsibility reports to attract more investment funds and external support [8]. Over time, good carbon performance eliminates the information asymmetry between internal and external investors, forming a unique competitive advantage. Additionally, a good carbon performance level helps industrial firms to establish a better brand reputation and public image, making them more favorable to consumers, as well as government funding support and policy assistance. Increased profits and accumulated funds can help companies to expand their production and operations, improve performance, and achieve long-term development.

4.3. Impact Analysis of Carbon Trading on Industrial Firm Value

In Table 6, the coefficient CET × Own and the coefficient CET are both significant at the level of 1%, indicating that ownership concentration contributes to the enhancing effect of the implementation of carbon-trading policy on firm value. The coefficient CET × Gov is significant negative at the level of 10%, indicating that government intervention will inhibit the enhancing effect of the implementation of carbon-trading policy on firm value. In addition, the coefficient CET × Own × Gov is negative and significant at the level of 1%, indicating that the synergy between internal and external firm governance inhibits the positive effect of carbon trading on firm value.

Table 6. Impact of carbon trading on firm value under internal and external governance.

	TQ
CET	0.202 ***
	(0.036)
CET × Own	0.006 ***
	(0.002)
CET × Gov	−1.226 *
	(0.661)
CET × Own × Gov	−0.164 ***
	(0.041)
Own × Gov	0.051 ***
	(0.015)
Gov	−1.355 **
	(0.670)
Own	−0.010 ***
	(0.001)
control	Yes
_cons	13.673 ***
	(0.675)
N	10,059
R^2	0.2876
year	yes
firm	yes
F	164.694

*: 10% significance level ($p < 0.10$); **: 5% significance level ($p < 0.05$); ***: 1% significance level ($p < 0.01$).

Firms participating in carbon emission trading policies tend to regulate their carbon emissions based on national and industry standards and are more inclined to demonstrate their good carbon information disclosure to society [43]. This not only improves the firm's image and changes stakeholders' attitudes toward the firm but also changes the capital market's favorability toward the firm, enhancing its value and creative ability [44]. Additionally, industrial firms participating in carbon trading have sufficient reasons to increase the value of their goods, better transferring carbon costs to consumers. Consumers often prefer to purchase environmentally friendly products, driving revenue for firms and thus increasing their value.

The ownership structure is a key factor influencing internal environmental governance within firms [45]. Firms with high ownership concentration typically mean that a few shareholders have significant influence and control [46], making it easier and more efficient for them to reach consensus and implement decisions [47]. High ownership concentration enables them to better formulate and implement long-term plans and strategies consistent with the overall interests of the firm. Such long-term planning not only helps to reduce the risks of firms under environmental regulatory policies [48], but also has the potential to benefit from policy changes.

A high level of government intervention in the location of the firm implies that the firm has more political connections [49]. The actual resource allocation of the firm may be subject to government influence, resulting in misallocation [50]. A weaker degree of government intervention means that firms have more autonomy in investment decisions in the environmental field [51]. Therefore, under strong government intervention, the implementation of carbon-trading policy has a negative effect on firm value.

4.4. Heterogeneity Analysis

4.4.1. Heterogeneity Analysis Based on Industry Attributes

According to the document issued by the China Securities Regulatory Commission in 2012 (Guidelines for the Industry Classification of Listed Enterprises), and the literature [41], coal mining and washing industry oil (B06); natural gas mining (B07); petroleum processing, coking, and nuclear fuel processing in manufacturing (C25), and electricity and heat, production and supply, and gas production and supply (D44 and D45) are defined as energy industries, while other industries (such as C14: food manufacturing, etc.) are defined as non-energy industries. Accordingly, this paper further tests the difference in industry heterogeneity on the change in firm value brought by the participation of industrial firms in carbon trading. The results are shown in Table 7.

Table 7. Heterogeneity based on industry attributes.

	Traditional Energy Firms TQ	**Non-Traditional Energy Firms TQ**
CET	0.218 ***	0.557 ***
	(0.074)	(0.038)
Control	Yes	Yes
_cons	7.473 ***	14.649 ***
	(1.786)	(0.661)
N	641	9418
R^2	0.3313	0.2281
year	yes	yes
firm	yes	yes
F	19.369	111.160

***: 1% significance level ($p < 0.01$).

The coefficients of non-energy industries are all significant and larger than those of energy industries, which indicates that the impacts of carbon trading on firm value varies by industries. Non-energy industries are generally easier to transition than energy industries, and typically involve relatively low technical barriers and relatively low transition costs. Moreover, non-energy industries tend to adopt more flexible production methods, and it is easier to adjust business modes and production processes to adapt to market changes and environmental protection requirements, so the effect of carbon trading on firm value is more pronounced in non-energy industries and less so in energy industries.

4.4.2. Heterogeneity Analysis Based on Industry Category

Industrial firms can be classified into mining, manufacturing, electricity, heat, gas, and water production and supply sectors. This section investigates the impact of carbon-trading policy on the valuation of different firm types. As shown in Table 8, the implementation of carbon-trading policy positively influences value enhancement across all three categories. Among them, the implementation of carbon-trading policy has the strongest effect on the value improvement of manufacturing firms.

Table 8. Heterogeneity based on industry category.

	Mining Firms	**Manufacturing Firms**	**Electricity, Heat, Gas, and Water Production and Supply Firms**
	TQ	**TQ**	**TQ**
CET	0.190 *	0.214 ***	0.157 ***
	(0.106)	(0.038)	(0.048)
Control	Yes	Yes	Yes
_cons	10.287 ***	14.168 ***	4.125 ***
	(2.376)	(0.774)	(1.307)
N	371	9219	469
R^2	0.4955	0.2984	0.4163
year	yes	yes	yes
firm	yes	yes	yes
F	19.030	205.754	16.527

*: 10% significance level ($p < 0.10$); ***: 1% significance level ($p < 0.01$).

The economic benefits of mining firms are mainly influenced by raw materials, labor, and equipment [52], while the carbon cost accounts for a small proportion of the benefit cost. Manufacturing firms are more capable of technological upgrading and innovation [53] and are likely to produce more competitive environmental products to meet market demand. As the products and services produced by the electricity, heat, gas, and water production and supply industries are basic, with high rigidity on the demand side and low price elasticity of demand, the product market has a strong trading stability [54], even if affected by the carbon-trading policy. In addition, these industries usually have relatively mature technologies and production facilities, which can relatively easily implement carbon reduction measures, such as improving energy efficiency and adopting clean energy. Therefore, it is easier for these industries to adapt to the requirements of carbon-trading policies and reduce carbon emissions than other industries.

4.4.3. Heterogeneity Analysis Based on Geographic Region

Due to there being no carbon-trading market in the northeast, China is divided into eastern, central, and western regions for heterogeneity analysis. We performed a group test, as shown in Table 9, and found that carbon trading increased the value of industrial firms in different regions. CET in the central region is positive, and at the level of 1% it is

significant. Compared to the eastern and western regions, the impact of carbon-trading policy on the central region's firm value is greater.

Table 9. Heterogeneity based on geographic region.

	Firms in Eastern China TQ	Firms in Central China TQ	Firms in West China TQ
CET	0.515 ***	0.719 ***	0.508 ***
	(0.044)	(0.089)	(0.157)
Control	Yes	Yes	Yes
_cons	14.403 ***	13.115 ***	14.876 ***
	(0.884)	(1.271)	(1.570)
N	6384	1695	1554
R^2	0.2399	0.2294	0.2248
year	yes	yes	yes
firm	yes	yes	yes
F	62.089	33.918	15.912

***: 1% significance level ($p < 0.01$).

In the central region of China, the development model is mainly focused on high-pollution and high-emission industries. Therefore, for these firms, participation in carbon trading can provide them with more room for value enhancement. The eastern firms themselves have a high degree of perfection, coupled with the region itself having achieved certain results in ecological governance, so there is little room for technological progress and firm value enhancement. In western China, technology and management are lagging behind, which has led to a slower response to the reduction effects of carbon-trading policies.

4.5. Robustness Test

4.5.1. PSM-DID Method

Before constructing the DID model, the PSM (Propensity Score Matching) method was used to avoid sample selection bias. The kernel density functions before and after PSM are shown in Figure 1, and the results of PSM-DID are shown in Table 10. It can be seen that carbon trading still significantly improves the value of industrial firms (the coefficient is 0.201, $p < 0.01$). This indicates that the carbon-trading system has a positive impact on the enhancement of industrial firm value, which is significant and continuous.

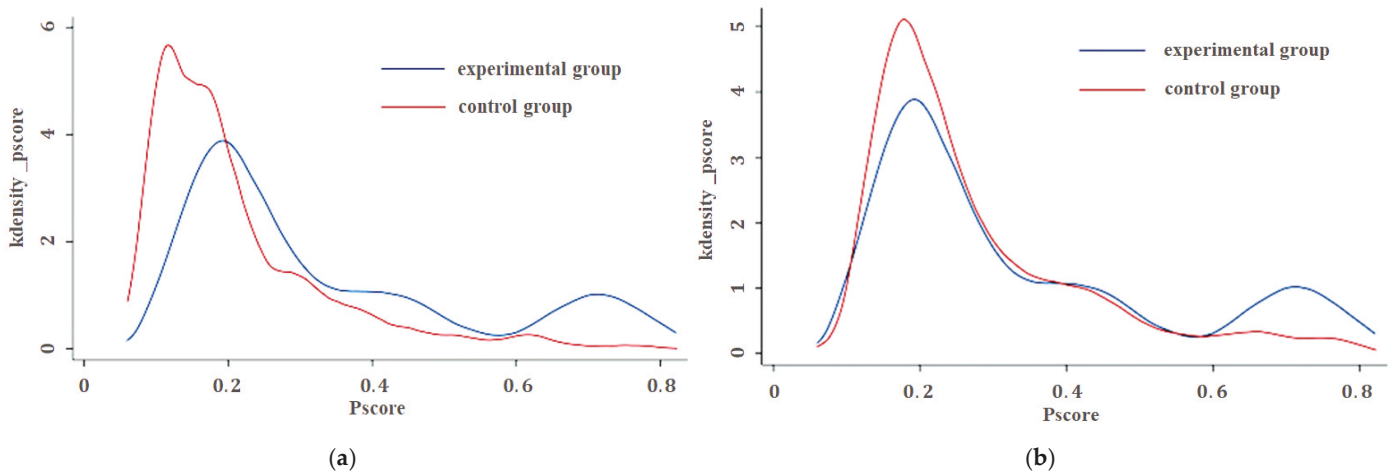


Figure 1. Kernel density function diagram. (**a**). Before PSM matching. (**b**). After PSM matching.

Table 10. Robustness test.

	PSM-DID TQ	Variable Substitution L.TQ
CET	0.201 ***	0.157 ***
	(0.035)	(0.041)
Control	Yes	Yes
_cons	13.971 ***	9.509 ***
	(0.562)	(0.651)
N	9962	8869
R^2	0.2894	0.2592
year	yes	yes
firm	yes	yes
F	212.762	170.545

***: 1% significance level ($p < 0.01$).

4.5.2. Method of Substitution Variables

Considering the possible lag effect of the impact of the implementation of carbon-trading policy on firm value, we replace the current period firm value with a first-order lag term. The results, shown in Column 2 of Table 10, are robust, and carbon trading also has a promoting effect on firm value with one lag period. This indicates that carbon-trading policy has a continuous improvement on firm value, and its impact is not limited to the current period of policy implementation, but can still have a positive impact in the subsequent period.

5. Conclusions

Driven by big data from the listed industrial firms, this study examined the direct effect of carbon trading on firm value, the indirect effect with firm carbon performance as the intermediary variable, and the synergy effect between internal and external governance of firms on carbon trading and firm value. Heterogeneity and robustness tests were also conducted. The results show the following.

(1) The implementation of carbon-trading policy can significantly improve the firm value. Carbon trading enhances a firm's value by improving its carbon performance. Ownership concentration strengthens the positive impact of carbon trading on firm value, while government intervention weakens this effect.
(2) Compared to energy firms, carbon trading has a more significant effect on the value of firms in non-energy industries. The policy promotes the improvement of firm value across three types of industries, with manufacturing firms experiencing the strongest effect. Additionally, the policy has a more pronounced impact on firm value in the central and western regions compared to the eastern region.

Based on the research conclusions, the following recommendations are proposed:

(1) To the government: When implementing carbon-trading policies, governments should reduce intervention levels, encourage market-driven carbon trading, and strengthen supervision to ensure fairness and transparency. In addition, efforts should be made to address regional disparities, particularly by filling gaps in the carbon-trading market in the northeast.
(2) To firms: Firms should increase investment in energy conservation and emission reduction, adopt more environmentally friendly production methods, and actively participate in carbon trading. Non-energy industry firms and those in the central and western regions should leverage carbon-trading policies to enhance firm value and seek cooperation with energy firms.

The current study has several key limitations and assumptions. First, it focuses on linear short-term effects, leaving the long-term impacts of carbon-trading policies unexplored. Second, it does not examine the interaction between carbon trading and other environmental policies, such as carbon taxes or green subsidies, which may influence outcomes. Third, the analysis lacks advanced quasi-experimental designs, such as DDD (difference-in-differences-in-differences) models, to isolate the specific effects of carbon trading from other overlapping policies. Finally, the study lacks an international comparative perspective, which could provide broader insights into how carbon trading functions across different regulatory and economic contexts.

In future work, we aim to investigate the long-term complex impacts of carbon-trading policies using dynamic panel data models and time-series analysis to better understand their persistent effects on firm value and carbon performance. Additionally, we plan to examine the interaction between carbon trading and other policies, such as carbon taxes and green subsidies, by employing policy interaction terms and agent-based simulations. Furthermore, we intend to utilize advanced quasi-experimental designs such as DDD models, to evaluate the combined effects of overlapping environmental policies. Finally, we will expand the scope of analysis to include international case studies, which will enable a comparative understanding of how carbon-trading policies function in different regulatory and economic contexts.

Author Contributions: Conceptualization, Z.P.; methodology, Z.P. and T.S.; software, Y.Z.; validation, Z.P., Y.Z. and T.S.; formal analysis, Z.P., Y.Z. and T.S.; investigation, T.S.; resources, T.S.; data curation, Z.P. and Y.Z.; writing—original draft preparation, Z.P. and Y.Z.; writing—review and editing, Z.P., Y.Z. and T.S.; visualization, Y.Z.; supervision, T.S.; project administration, Z.P.; funding acquisition, Z.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Science Foundation of China (NSFC), grant number 72271033; International (Regional) Cooperation and Exchange (ICE) Projects of the National Natural Science Foundation of China (NSFC), grant number 72311540159, and the Central University Basic Research Fund of China, grant number 2-9-2022-021. And the APC was funded by 72271033.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors thank the Natural Science Foundation of China (NSFC) [grant number 72271033].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yang, W.; Pan, Y.; Ma, J.; Yang, T.; Ke, X. Effects of allowance allocation rules on green technology investment and product pricing under the cap-and-trade mechanism. *Energy Policy* **2020**, *139*, 111333. [CrossRef]
2. Chen, S.; Shi, A.; Wang, X. Carbon emission curbing effects and influencing mechanisms of China's Emission Trading Scheme: The mediating roles of technique effect, composition effect and allocation effect. *J. Clean. Prod.* **2020**, *264*, 121700. [CrossRef]
3. Zhang, W.; Li, G.; Guo, F. Does carbon emissions trading promote green technology innovation in China? *Appl. Energy* **2022**, *315*, 119012. [CrossRef]
4. Zhang, Y.-J.; Liang, T.; Jin, Y.-L.; Shen, B. The impact of carbon trading on economic output and carbon emissions reduction in China's industrial sectors. *Appl. Energy* **2020**, *260*, 114290. [CrossRef]
5. Guo, Q.; Su, Z.; Chiao, C. Carbon emissions trading policy, carbon finance, and carbon emissions reduction: Evidence from a quasi-natural experiment in China. *Econ. Change Restruct.* **2022**, *55*, 1445–1480. [CrossRef]
6. Yang, S.; Lu, T.; Huang, T.; Wang, C. Re-examining the effect of carbon emission trading policy on improving the green innovation of China's enterprises. *Environ. Sci. Pollut. Res.* **2023**, *30*, 7696–7717. [CrossRef]

7. Zhang, Y.; Liu, J. Does carbon emissions trading affect the financial performance of high energy-consuming firms in China? *Nat. Hazards* **2019**, *95*, 91–111. [CrossRef]
8. Benkraiem, R.; Shuwaikh, F.; Lakhal, F.; Guizani, A. Carbon performance and firm value of the World's most sustainable companies. *Econ. Model.* **2022**, *116*, 106002. [CrossRef]
9. Yang, Z.; Yuan, Y.; Zhang, Q. Carbon emission trading scheme, carbon emissions reduction and spatial spillover effects: Quasi-experimental evidence from China. *Front. Environ. Sci.* **2021**, *9*, 824298. [CrossRef]
10. Peng, H.; Qi, S.; Cui, J. The environmental and economic effects of the carbon emissions trading scheme in China: The role of alternative allowance allocation. *Sustain. Prod. Consum.* **2021**, *28*, 105–115. [CrossRef]
11. Al Rabab'A, E.A.-F.; Rashid, A.; Shams, S. Corporate carbon performance and cost of debt: Evidence from Asia-Pacific countries. *Int. Rev. Financ. Anal.* **2023**, *88*, 102641.
12. Shi, W.; Sang, J.; Zhou, J.; Ding, X.; Li, Z. Can carbon emission trading improve carbon emission performance? Evidence from a quasi-natural experiment in China. *Environ. Sci. Pollut. Res.* **2023**, *30*, 124028–124040. [CrossRef] [PubMed]
13. Yu, X.; Chen, H.; Li, Y. Impact of carbon emission trading mechanism on carbon performance based on synthetic control method. *China Popul. Resour. Environ.* **2021**, *31*, 51–61.
14. Zhang, S.; Cheng, L.; Ren, Y.; Yao, Y. Effects of carbon emission trading system on corporate green total factor productivity: Does environmental regulation play a role of green blessing? *Environ. Res.* **2024**, *248*, 118295. [CrossRef]
15. He, M.; Zhu, X.; Li, H. How does carbon emissions trading scheme affect steel enterprises' pollution control performance? A quasi natural experiment from China. *Sci. Total Environ.* **2022**, *858*, 159871. [CrossRef]
16. Yu, X.; Shi, J.; Wan, K.; Chang, T. Carbon trading market policies and corporate environmental performance in China. *J. Clean. Prod.* **2022**, *371*, 133683. [CrossRef]
17. Ismanu, S.; Kusmintarti, A.; Sulistiono, S. Implementation of enterprise risk management and determinants of company performance. *Acad. Strateg. Manag. J.* **2022**, *21*, 1–8.
18. Hao, X.; Chen, F.; Chen, Z. Does green innovation increase enterprise value? *Bus. Strategy Environ.* **2021**, *31*, 685–1247. [CrossRef]
19. Han, M.; Huang, W. The impact of China's carbon emission allowance price on the value of energy enterprises. *J. China Univ. Pet. (Ed. Soc. Sci.)* **2022**, *38*, 20–28. [CrossRef]
20. Lin, P.; Lin, M.; Lin, B. Carbon emission trading system and firm value under carbon peaking and carbon neutrality. *Account. Econ. Res.* **2023**, *37*, 135–147. [CrossRef]
21. Niu, X.; Zhang, Y.; Li, B.; Chen, Z.; Ni, G.; Lyu, N. How does carbon emission trading scheme affect enterprise market value? A roadmap towards natural resources sustainability. *Resour. Policy* **2024**, *88*, 104542. [CrossRef]
22. Yuan, M. Will regulated firms benefit from carbon emission trading system? Evidence from a market power perspective. *Environ. Sci. Pollut. Res.* **2023**, *30*, 103001–103016. [CrossRef] [PubMed]
23. Ren, X.; Ma, Q.; Sun, S.; Ren, X.; Yan, G. Can China's carbon trading policy improve the profitability of polluting firms: A retest of Porter's hypothesis? *Environ. Sci. Pollut. Res.* **2023**, *30*, 32894–32912.
24. Qi, Y.; Yuan, M.; Bai, T. Where will corporate capital flow to? Revisiting the impact of China's pilot carbon emission trading system on investment. *J. Environ. Manag.* **2023**, *336*, 117671. [CrossRef] [PubMed]
25. Du, H.; Zhang, Y.; Evans, O.-M.; Chen, W. Will emission trading promote enterprise diversification? Evidence from China. *Environ. Sci. Pollut. Res.* **2023**, *30*, 79194–79214. [CrossRef]
26. Gao, L.; Jiang, C.X.; Mekhaimer, M. Count on subordinate executives: Internal governance and innovation. *J. Bank. Financ.* **2023**, *154*, 106931. [CrossRef]
27. Xu, W.; Jiang, X. Does the D&O insurance enhance firm value of listed companies in China. *J. Appl. Stat. Manag.* **2023**, *42*, 158–174.
28. Qian, X.; Ding, H.; Ding, Z. Governmental inspection and firm environmental protection expenditure: Evidence from China. *Econ. Model.* **2023**, *123*, 106284. [CrossRef]
29. Li, M.; Liu, H.; Chiang, Y. Government intervention, leverage adjustment, and firm performance: Evidence from defaulting firms. *Pac.-Basin Financ. J.* **2022**, *76*, 101885. [CrossRef]
30. Gu, D.; Wang, Q.; Chai, Y.; Yang, X.; Zhao, W.; Li, M.; Zolotarev, O.; Xu, Z.; Zhang, G. Identifying the risk factors of allergic rhinitis based on zhihu comment data using a topic-enhanced word-embedding model: Mixed method study and cluster analysis. *J. Med. Internet Res.* **2024**, *26*, e48324. [CrossRef]
31. Zhao, W.; Gu, D.; Yang, X.; Jia, M.; Liang, C.; Wang, X.; Zolotarev, O. MedT2T: An adaptive pointer constrain generating method for a new medical text-to-table task. *Future Gener. Comput. Syst.* **2024**, *161*, 586–600. [CrossRef]
32. Qin, X.; Hou, Y. Smart city construction, industrial agglomeration, and total factor productivity of enterprises. *J. Nanjing Univ. Financ. Econ.* **2023**, *6*, 99–110.
33. Tang, Q.; Xu, X.; Cao, Y. Stock right incentive, research investment and sustainable development of enterprises evidence from Chinese listed companies. *J. Shanxi Univ. Financ. Econ.* **2009**, *31*, 77–84.
34. Han, Y.; Li, Z.; You, J.; Dong, H. Policy effect and mechanism of new urbanization on urban carbon balance. *Urban Probl.* **2024**, *1*, 16–26. [CrossRef]

35. Wang, H.; Li, S.; Wang, Y.; Gong, Y. Evaluation of scientific and technological innovation efficiency of national specialized and sophisticated "little giant" firms and its comparison of five urban agglomerations. *Sci. Technol. Manag. Res.* **2023**, *43*, 65–74.
36. Ma, J.; Qin, F.; Xie, X. Decoupling the inbound tourism carbon emissions off the tourism economic growth. *J. Xinjiang Univ. (Philos. Soc. Sci.)* **2019**, *47*, 16–23. [CrossRef]
37. Jin, Y.; Chen, Z.; Lu, M. Industry agglomeration in China: Economic geography, new economic geography and policy. *Econ. Res. J.* **2006**, *4*, 79–89.
38. Wu, F.; Hu, H.; Lin, H.; Ren, X. Enterprise digital transformation and capital market performance: Empirical evidence from stock liquidity. *J. Manag. World* **2021**, *37*, 130–144+10. [CrossRef]
39. Chen, S.; Wang, Y.; Albitar, K.; Huang, Z. Does ownership concentration affect corporate environmental responsibility engagement? The mediating role of corporate leverage. *Borsa Istanb. Rev.* **2021**, *21*, S13–S24. [CrossRef]
40. Chen, G. Land resource misallocation, industrial structure and haze pollution: An empirical study based on spatial econometrics and dynamic panel threshold model. *China Soft Sci.* **2022**, *12*, 143–152.
41. Liu, X.; Zhang, X. Industrial agglomeration, technological innovation and carbon productivity: Evidence from China. *Resour. Conserv. Recycl.* **2021**, *166*, 105330. [CrossRef]
42. Liu, N.; Fan, L.; Li, S.; Chen, X.; Mao, L. The Influence of the carbon trading system on the enterprise' carbon emission reduction performance. *Syst. Eng.* **2022**, *40*, 13–23.
43. Ren, X.; Ma, Q.; Liu, Y.; Zhao, G. The impact of carbon trading policy on the economic performance of highly polluting industrial enterprises: Empirical analysis based on multiple mediating effect model. *Resour. Sci.* **2020**, *42*, 1750–1763.
44. Jiang, Y.; Luo, L.; Xu, J.; Shao, X. The value relevance of corporate voluntary carbon disclosure: Evidence from the United States and BRIC countries. *J. Contemp. Account. Econ.* **2021**, *17*, 100279. [CrossRef]
45. Puangyanee, S.; Yaowapanee, P.; Duangsawang, K.; Jermsittiparsert, K. The Influence of the shareholding structure on economic performance through good corporate governance of listed companies in the stock exchange of Thailand. *Int. J. Innov. Creat. Change* **2019**, *7*, 116–133.
46. Xu, W.; Zhang, R.; Liu, Y.; Liu, P. Classification governance, controlling party governance mechanism and innovation bonus release: A study based on the analysis of state-controlled listed companies. *Nankai Bus. Rev.* **2018**, *21*, 90–102.
47. Garritty, C.; Hamel, C.; Hersi, M.; Butler, C.; Monfaredi, Z.; Stevens, A.; Nussbaumer-Streit, B.; Cheng, W.; Moher, D. Assessing how information is packaged in rapid reviews for policy-makers and other stakeholders: A cross-sectional study. *Health Res. Policy Syst.* **2020**, *18*, 112. [CrossRef]
48. Xu, Y.; Li, S.; Zhou, X.; Shahzad, U.; Zhao, X. How environmental regulations affect the development of green finance: Recent evidence from polluting firms in China. *Renew. Energy* **2022**, *189*, 917–926. [CrossRef]
49. Yang, L.; Zhang, J. Political connections, government intervention and acquirer performance in cross-border mergers and acquisitions: An empirical analysis based on Chinese acquirers. *World Econ.* **2015**, *38*, 1505–1525. [CrossRef]
50. Zheng, W. Research on the promotion effect of smart city construction on high-quality employment based on the empirical evidence of multi-period DID. *Guizhou Soc. Sci.* **2023**, *7*, 128–136. [CrossRef]
51. Su, T.; Meng, L.; Zhang, J. China's carbon market pilot and enterprise green transition: An analysis of effects and mechanism. *RD Manag.* **2022**, *34*, 81–96. [CrossRef]
52. Lu, H.; Wu, Z. Relationship between energy-consuming right trading system and low-carbon transformation of energy consumption structure. *Resour. Sci.* **2023**, *45*, 1181–1195. [CrossRef]
53. Chen, H.; Zhong, T.; Lee, J. Capacity reduction pressure, financing constraints, and enterprise sustainable innovation investment: Evidence from Chinese manufacturing companies. *Sustainability* **2020**, *12*, 10472. [CrossRef]
54. Xu, H.; Hou, Z.; Jiang, Y.; Liu, W. Analysis on the discount for low liquidity in China's stock market evidence from the new third board. *China Econ. Stud.* **2022**, *5*, 103–118. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

How to Self-Disclose? The Impact of Patients' Linguistic Features on Doctors' Service Quality in Online Health Communities

Mengyuan Peng [1,†], Kaixuan Zhu [1,†], Yadi Gu [2,†], Xuejie Yang [3,*], Kaixiang Su [1,†] and Dongxiao Gu [1,4,*]

[1] School of Management, Hefei University of Technology, Hefei 230009, China; 2022110891@mail.hfut.edu.cn (M.P.); kaixuan_zhu@mail.hfut.edu.cn (K.Z.); kaixiang_su@mail.hfut.edu.cn (K.S.)
[2] Mental Health Education Center, University of Shanghai for Science and Technology, Shanghai 200093, China; yadigu2002@163.com
[3] School of Humanity and Law, Hefei University of Technology, Hefei 230009, China
[4] Laboratory of Data Science and Smart Society Governance of the Ministry of Education, Hefei University of Technology, Hefei 230009, China
* Correspondence: xuejie_y@126.com (X.Y.); dongxiaogu@yeah.net (D.G.)
† These authors contributed equally to this work.

Abstract: In online medical consultations, patients convey their medical condition through self-disclosure, and the linguistic features of this disclosure, as signals, may significantly impact doctors' diagnostic behavior and service quality. Based on signaling theory, this paper collects consultation data from a large online medical platform in China, employs text mining and classification techniques to extract relevant variables, and applies econometric models to empirically examine the effect of patients' self-disclosure linguistic features on the quality of online medical services. The results indicate that the completeness and readability of patients' self-disclosure have a significant positive impact on the quality of doctors' services, while the expertise and positive sentiment of the disclosure have a significant negative effect. From the perspective of signaling theory, this study reveals the mechanism through which patients' self-disclosure linguistic features influence doctors' online consultation behavior, providing an important theoretical foundation for promoting online doctor–patient interaction and enhancing patient well-being.

Keywords: self-disclosure; linguistic features; service quality; informational support; emotional support

1. Introduction

In recent years, online medical consultation platforms have gradually become a crucial channel for patients to access healthcare services, overcoming the temporal and spatial constraints of traditional offline healthcare and greatly facilitating doctor–patient communication [1]. This not only optimizes the allocation of medical resources but also enhances the accessibility and equity of healthcare services, which is particularly significant for patients in remote areas [2]. In China, an emerging economy characterized by significant urban–rural disparities and cultural diversity, the structural imbalance in healthcare resources combined with the influence of traditional cultural beliefs creates a unique context for the development of online medical consultations [3,4]. Specifically, although tertiary-A hospitals (the highest-level medical institutions) account for only 7.6% of all hospitals nationwide, they handle nearly 47% of outpatient visits. The absence of an effective hierarchical diagnosis and treatment system has resulted in the concentration of high-quality healthcare

services in major cities, where these resources are persistently overburdened [5,6]. Given these challenges, online medical consultations have emerged as a vital means of mitigating the structural imbalance in healthcare resources. However, unlike offline consultations, online consultations lack direct, in-person interaction, where patients and doctors can convey information through non-verbal signals such as eye contact, facial expressions, and body language [7]. Furthermore, traditional diagnostic methods like palpation and auscultation cannot be performed in the online environment. In this context, language itself becomes the primary medium of communication between doctors and patients. Therefore, the linguistic expression of patients plays a pivotal role in facilitating doctor–patient communication and enhancing doctors' engagement in online consultations.

The primary goal of patients in seeking medical consultations is to obtain accurate diagnoses and effective treatment recommendations, as well as a satisfactory healthcare experience. Doctors' online engagement and service quality are critical variables that determine patients' trust and satisfaction, playing a significant role in facilitating doctor–patient communication and enhancing patient well-being [8,9]. Existing studies typically assess the quality of offline healthcare services across three dimensions: environmental quality, outcome quality, and interaction quality [10]. However, the computer-mediated communication (CMC) nature of online healthcare introduces two structural differences. First, the absence of a physical environment renders traditional assessments of environmental quality inapplicable [11]. Second, online consultations primarily focus on information exchange and lack physical treatment interventions, making outcome quality more difficult to measure [12]. Given these constraints, interaction quality has become the core dimension for evaluating service quality in online healthcare. Current research generally uses two types of observable indicators to measure interaction quality in online consultations. The first type involves behavioral metrics such as doctors' response speed, and frequency of interaction, which are used to quantify doctors' service effort [13–15]. The second focuses on content analysis, assessing the informational support and emotional support provided by doctors based on the textual content of their communication [16]. In addition, subjective evaluations from patients, such as doctor ratings and online reviews, are commonly used to measure perceived service quality, although these assessments typically lag behind objective behavioral and content-based indicators [17,18].

How can doctors be encouraged to provide higher-quality services in online consultations? Given the importance of doctors' engagement, existing research has explored the motivation and influencing factors behind doctors' contributions in online consultations, from perspectives such as platform incentives, doctors' traits, and reputation [19–21]. However, even for patients with similar conditions, the quality of diagnosis and service provided by the same doctor often varies. Therefore, exploring the influence of patient-related factors on doctors' consultation behavior from the patient's perspective holds significant theoretical and practical implications. Existing research has considered various factors from the patient's perspective that influence doctors' behavior, such as monetary and non-monetary incentives. However, the linguistic expression of patients during consultations, particularly the manner and characteristics of their self-disclosure, has not received sufficient attention [14,22]. While a few studies have examined the impact of patients' self-disclosure on doctors' consultation behavior, most have focused on the depth or content of the disclosure, lacking a detailed and in-depth analysis from a linguistic perspective [23].

Language shapes not only the content of communication (what is said) but also its form and style (how it is said) [24]. From the perspective of signaling theory, patients can adjust their linguistic signals to influence doctors' responses and obtain higher-quality services and medical advice. Although individuals are often unaware of consciously controlling their language style, these linguistic features can significantly affect how infor-

mation recipients perceive and respond to the message [25]. Therefore, based on signaling theory, this study constructs an empirical model using real consultation data from the platform, Haodf.com (accessed on 15 October 2024), to examine the impact of the linguistic characteristics of patients' self-disclosure on the quality of doctors' services in online consultations. Specifically, the study focuses on four core linguistic features of patient self-disclosure: (1) completeness, (2) readability, (3) expertise, and (4) sentiment valence. These features were selected for two main reasons. First, they collectively involve the four key dimensions of textual linguistic analysis: lexical, syntactic, semantic, and pragmatic [26]. The use of medical terminology reflects the patient's medical expertise and pertains to the lexical dimension [27]; readability primarily involves syntactic structure [28]; sentiment valence belongs to the semantic dimension [29]; and completeness reflects the extent and adequacy of the information disclosed by patients, involving pragmatic as well as other linguistic dimensions [30]. Second, these linguistic features are closely related to the context of doctor–patient communication. From the perspective of signaling theory, these features enhance the observability of patients' signals, which can influence the perceived psychological distance between doctors and patients. Specifically, completeness enables doctors to fully understand the patient's condition; high readability facilitates quick and accurate comprehension of patient concerns; expertise reflects patients' health literacy and may reduce communication barriers; and emotional expression directly affects emotional resonance and the quality of interaction between doctors and patients [31,32].

In summary, this study takes a patient-centered perspective to further advance research in the field of online doctor–patient interactions. By examining how linguistic features of patient disclosure influence doctors' quality, the study offers theoretical contributions and practical implications for improving online healthcare services.

2. Theoretical Foundation and Literature Review

2.1. Signaling Theory

Signaling theory was first proposed by Spence in 1973, primarily to explain the communication and transmission of information between information seekers and providers in situations of information asymmetry [33]. Initially, the theory was used to explain how sellers send signals to buyers to promote consumer purchasing behavior in the context of information asymmetry [34,35]. Nowadays, it has been widely applied in studies on online social platforms. For example, Choi et al. employed signaling theory as a theoretical model to examine the impact of product quality cues on the sales of digital video games [36]. Benlian et al. found that information technology features serve as important signals affecting user trust and engagement in online communities [37]. In the context of online consultation platforms, signaling theory provides a framework to understand how patients use external cues (signals) to convey their health conditions and other relevant information to doctors, thereby facilitating doctor–patient interaction and communication [4]. Patients and doctors are two parties with a high degree of information asymmetry: patients possess private information about their health condition but lack medical expertise, while doctors have medical knowledge and skills but are unaware of the specific details of the patient's condition [38]. In such an information asymmetry context, patients may not only provide specific details of their health condition to doctors but may also adjust their method of self-disclosure, that is, linguistic signals, to receive more attention and service from the doctor [39]. In patient-to-patient communities, signaling theory has been applied to examine how the linguistic characteristics of posts influence peer attention and support [30,40]. Similarly, in online doctor–patient interactions, language serves not only as a medium for conveying medical information but also as a critical signal shaping the doctors' responses. Khurana et al. identified doctors' online responses as signals that

significantly impact patient satisfaction [41]. Moreover, Zhang et al. demonstrated that patient-generated information serves as an essential linguistic signal, influencing subsequent patients' information-seeking behavior and healthcare choices [42]. Thus, this study applies signaling theory to investigate how patients' linguistic features serve as signals that shape doctors' service quality and doctor–patient interactions.

2.2. Linguistic Characteristics of Self-Disclosure

Self-disclosure refers to the intentional sharing of personal information with others and is an important factor in the development of interpersonal relationships [43]. On social media platforms, individuals enhance familiarity and a sense of connection through self-disclosure, thereby gaining more attention and interaction [44]. Furthermore, self-disclosure helps to build trust and facilitates cooperation and the achievement of transactions [45]. Previous studies have explored various aspects of self-disclosure in social media. For example, Quirdi investigated the predictors of inappropriate and career-oriented self-disclosure [46]. Bae et al. examined the relationship between self-disclosure of different types of background information and social support [47]. However, most existing studies have focused primarily on the level and content of self-disclosure, with limited attention paid to the systematic analysis of its linguistic features [43,48].

Linguistic features serve not only as a means of conveying the communicator's intentions but also have the potential to evoke behavioral changes in the recipient through various performative functions [49,50]. The linguistic characteristics of a text typically cover four dimensions: vocabulary, syntax, semantics, and pragmatics [26]. The vocabulary dimension focuses on the use of words and phrases, such as vocabulary richness and the use of technical terms. In the healthcare domain, the use of health-related terminology reflects the patient's level of expertise [51,52]. The syntax dimension examines sentence structure complexity and the coherence between sentences, which directly affect the readability of the text [28]. The semantic dimension investigates the meanings conveyed by the language, with sentiment valence being a key semantic feature [29,53]. The pragmatic dimension emphasizes the social functions of language, where the length of the text reflects the level of completeness in the expression [54]. In the field of doctor–patient interaction, existing studies have primarily focused on the linguistic characteristics of doctors, such as emotions, topic diversity, and vocabulary richness, exploring their impact on patient feedback [55,56]. In contrast, research on the linguistic characteristics of patients mainly focuses on patient communities, with relatively few studies on doctor–patient interaction scenarios [30,40]. Moreover, existing research often emphasizes common, single features such as emotions or text length, lacking a systematic and fine-grained analysis specific to the context of online consultations [23,57,58]. Therefore, this study, grounded in the context of online consultations and the four dimensions of linguistic features, explores the completeness, readability, expertise, and sentiment polarity of patients' self-disclosure, and their potential impact on the quality of doctor services.

2.3. Doctor's Consultation Behavior and Service Quality

In online consultations, doctors' behaviors influence the patient's healthcare experience as well as subsequent treatment decisions [59]. Social support from doctors is a crucial indicator of the quality of doctor–patient consultations, primarily in the forms of informational support and emotional support. Informational support helps patients better understand their condition, reduce cognitive uncertainty, and make more informed treatment choices [60,61], while emotional support alleviates patients' psychological burden, enhances their confidence in treatment, and thereby facilitates recovery [16,62]. To date, research on doctors' online consultation behaviors has primarily focused on their impact

on patient satisfaction and subsequent decision-making [18,63,64]. For instance, Yang et al. investigated how doctors' interaction rates and response times in online consultations influence patient satisfaction [13], while Khurana et al. found that doctors' responses to patients' inquiries positively affect patients' recommendations [41]. In contrast, studies exploring the motivations and influencing factors behind doctors' online behaviors remain limited, with most focusing on doctors' personal characteristics, status, incentives, and online reviews [4,19,20]. Communication in online consultations primarily occurs through language, which serves as the primary means by which patients convey to doctors. However, existing research has overlooked the impact of patients' linguistic expressions on doctors' consultation behaviors, focusing instead on other aspects of communication. Therefore, this study aims to examine doctors' consultation behaviors and service quality from the perspective of patients' self-disclosure linguistic features.

3. Research Hypotheses and Model

3.1. Research Hypotheses

This study focuses on the context of online doctor–patient interactions, exploring the impact of patients' self-disclosure on doctors' consultation behaviors and service quality based on four linguistic dimensions (vocabulary, syntax, semantics, and pragmatics). Specifically, it examines how the completeness, readability, expertise, and sentiment polarity of patients' self-disclosure influence doctors' behaviors and service quality.

3.1.1. Completeness

Completeness refers to the level of detail in patients' self-disclosure, reflecting the adequacy of the information provided. Generally, longer texts contain more self-disclosure, thus conveying a richer amount of information. Previous studies have shown that individuals who disclose more information are more likely to receive social support [23]. In the context of electronic healthcare, detailed information helps doctors gain a more comprehensive understanding of the patient's condition, thereby enhancing doctor–patient interaction and response [65]. Specifically, a thorough and complete description of the patient's health status not only facilitates more accurate diagnoses and treatment decisions but also conveys the patient's concern about their condition and expectation for assistance, prompting more attention and positive responses from the doctor. Thus, we propose the following hypotheses:

H1a. *The completeness of patients' self-disclosure positively influences the informational support provided by doctors in online consultations.*

H1b. *The completeness of patients' self-disclosure positively influences the emotional support provided by doctors in online consultations.*

3.1.2. Readability

Readability refers to the ease with which a text can be understood, also known as comprehensibility [66]. Readability is typically negatively correlated with text perplexity: higher perplexity usually indicates more complex language structures, grammar, and vocabulary choices, resulting in lower readability [40]. In online consultations, the readability of patients' self-disclosure directly impacts how well doctors can understand and respond to the information. Patients' communication abilities vary, and those with stronger expression skills tend to describe their health condition and needs more clearly and concisely, enabling doctors to quickly comprehend the situation and respond accurately. In contrast, texts with poor readability increase the cognitive load and comprehension cost for doctors, reducing

communication efficiency and limiting their ability to provide assistance. Therefore, we propose the following hypotheses:

H2a. *The readability of patients' self-disclosure positively influences the informational support provided by doctors in online consultations.*

H2b. *The readability of patients' self-disclosure positively influences the emotional support provided by doctors in online consultations.*

3.1.3. Expertise

Expertise refers to the knowledge and skills demonstrated by patients in the healthcare field, often closely associated with health literacy. It can be understood as an individual's ability to acquire, comprehend, and apply health information, which is crucial for making informed health decisions and effectively managing healthcare needs [67]. Expertise is a key factor influencing doctor–patient communication and health outcomes [68]. During consultations, patients who use specialized terms can convey their understanding of medical knowledge and their condition more clearly, reducing communication barriers and misunderstandings. Furthermore, according to communication accommodation theory, using vocabulary and language that matches the communication partner's style significantly enhances communication effectiveness and acceptance [26]. This professional expression style aligns more closely with the doctor's communication, encouraging more interaction. Therefore, we propose the following hypotheses:

H3a. *The expertise of patients' self-disclosure positively influences the informational support provided by doctors in online consultations.*

H3b. *The expertise of patients' self-disclosure positively influences the emotional support provided by doctors in online consultations.*

3.1.4. Sentiment Valence

Self-disclosure often involves sentiment valence [69]. Previous studies have shown that negative emotions significantly impact communication, with recipients typically responding more strongly to negative information [70,71]. In online social media, negative emotions are often more contagious and attract more attention than positive ones [72]. In online consultations, the sentiment polarity conveyed by patients sends important signals to doctors. Negative emotions tend to draw more attention from doctors, promoting more support and assistance, especially when emotional responses are strong. In such cases, doctors may perceive that the patient needs more care and intervention. Thus, we hypothesize the following:

H4a. *Positive sentiment valence in patients' self-disclosure negatively influences the informational support provided by doctors in online consultations.*

H4b. *Positive sentiment valence in patients' self-disclosure negatively influences the emotional support provided by doctors in online consultations.*

3.2. Research Model

In summary, based on signal theory, this study analyzes and tests the impact of the linguistic features of patients' self-disclosure on doctors' service quality. The overall research model is shown in Figure 1.

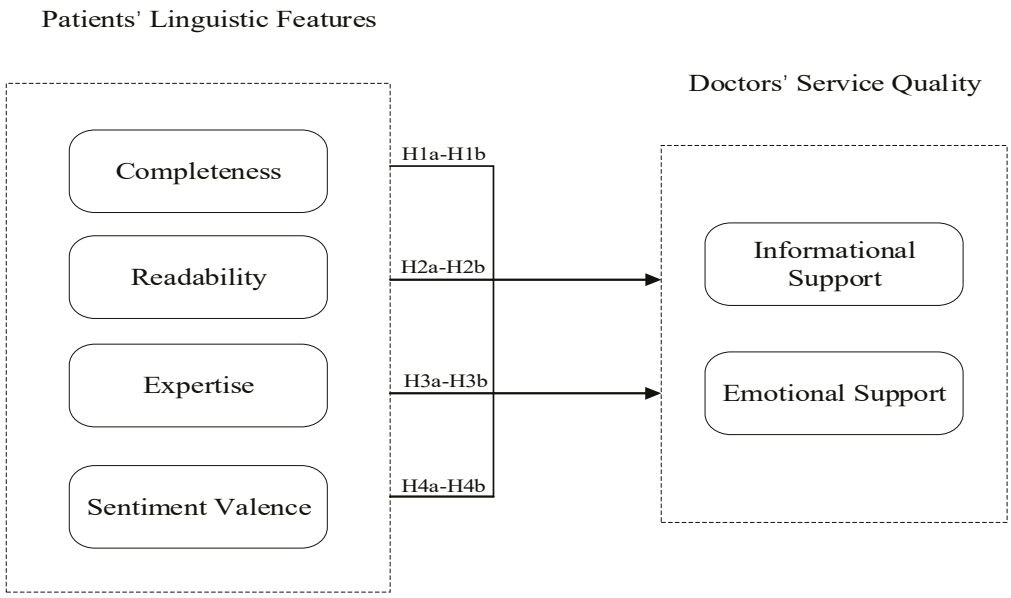


Figure 1. Research model.

4. Research Methodology

4.1. Research Context and Data Collection

China, the world's most populous country, faces persistent shortages and structural disparities in healthcare resources [73]. The absence of a strict hierarchical medical system, coupled with weak primary care and low patient trust, has driven excessive reliance on specialist care, even for minor conditions [74]. This overconcentration of resources has intensified healthcare access difficulties, making it a major public health challenge [75]. In this context, the rapid development of internet-based healthcare has created new pathways for restructuring doctor–patient interactions. This study examines how patients' language expression strategies in the Chinese-language context influence the quality of online medical consultations. The findings not only inform efforts to improve online healthcare but also provide insights for cross-cultural digital health research.

This study employs a Python web scraping tool (Python version 3.7.13) to collect data from Haodf.com, a major online healthcare platform in China, which provides extensive doctor profiles and doctor–patient consultation records, making it an ideal context for this research [76]. We collected all consultation records and relevant doctor profiles for two diseases—gastritis and liver cancer—between June 2014 and September 2023. A total of 67,871 doctor–patient consultation records were collected. After filtering for text and image-based consultation data and performing data cleaning, 52,730 valid records were obtained from 4871 doctors, including 26,654 records related to gastritis and 26,076 records related to liver cancer.

4.2. Variables

4.2.1. Dependent Variables

Information Support: The amount of information provided by the doctor during a round of doctor–patient interaction, measured by the number of content words. Content words refer to words that have concrete meanings, can independently serve as sentence

components, and carry lexical and grammatical significance, including nouns, verbs, adjectives, adverbs, numerals, classifiers, pronouns, state words, and differentiating words [28]. In this study, text segmentation and part-of-speech tagging were performed using the Jieba segmentation tool, with analysis based on a medical terminology dictionary. The number of content words was then calculated.

Emotional Support: The total amount of emotional support provided by the doctor during a round of doctor–patient interaction [29]. A BERT classifier was used to train a text classification model to determine whether emotional support was provided in each doctor's response (labeled as 0 for no emotional support and 1 for emotional support) [77]. The sum of emotional support provided during a round of interaction was then calculated. The BERT classifier achieved a 96% accuracy on the test set. The detailed text classification process is provided in Appendix A.1.

4.2.2. Independent Variables

Completeness: The degree of detail in patients' self-disclosure during online consultations, measured by the total length of the patient's text in a round of doctor–patient interaction.

Readability: The ease of understanding the text. Based on existing research, readability is measured by 1 minus the normalized perplexity of the sentence [40]. Perplexity is an indicator used to assess a language model's ability to predict text, representing the model's 'uncertainty' or 'complexity' in relation to the text. The lower the perplexity of the sentence, the simpler the text structure and the more natural the grammar, making it easier to understand. This study used a Chinese pre-trained GPT-2 model to calculate the perplexity of the patient's text.

Expertise: The average number of medical terms used by the patient during a round of doctor–patient consultation, reflecting the patient's expertise. This study used the medical terminology dictionary from Tsinghua University Open Chinese Thesaurus and supplemented it with terms related to stomach and liver diseasesbased on the content of the doctor–patient dialogue. The average number of medical terms in the patient's text was calculated using a Python program [78].

Sentiment Valence: The average emotional tendency expressed by the patient during a round of doctor–patient interaction. A BERT model was used for text classification, categorizing each patient's consultation text into three sentiment categories: negative (-1), neutral (0), and positive (1) [71]. The average sentiment valence for the patient's text in a round of interaction was then calculated. A value closer to 1 indicates more positive sentiment, while a value closer to -1 indicates more negative sentiment. The text classification method used is the same as that for emotional support, with a classification accuracy of 90% on the test dataset. The detailed text classification process is also shown in Appendix A.1.

4.2.3. Control Variables

To control for potential confounding factors, we introduced a series of control variables, including patient gender, patient age, doctor gender, doctor's employment grade, doctor's recommendation heat, hospital level, and communication time. The descriptions of all variables are shown in Table 1.

Table 1. Description of variables.

Variable Type	Variable	Description
Dependent Variables	Emotional Support (DEmoS)	The total amount of emotional support expressed by doctors in a round of doctor–patient interaction.
	Informational Support (DInfoS)	The total amount of information support expressed by the doctor in a round of doctor–patient interaction.
Independent Variable	Completeness	The degree of detail in patients' self-disclosure is measured by the total length of the patient's text in a round of doctor–patient interaction.
	Readability	The comprehensibility of patient text in a round of doctor–patient interaction.
	Expertise	The average number of medical terms used by patients in a round of doctor–patient interactions.
	Sentiment Polarity (Psentiment)	Average patient emotional polarity during a doctor–patient interaction. The closer it is to 1, the more positive the emotion, and the closer it is to −1, the more negative the emotion.
Control Variables	Patient Gender (PGender)	0 represents male, and 1 represents female.
	Patient Age (PAge)	The actual age of the patient, excluding patients with age 0.
	Doctor Gender (DGender)	0 represents male, and 1 represents female.
	Employment Grade (DEmGrade)	Categorical variable, doctor's professional title: 1 for physician, 2 for attending physician, 3 for associate chief physician, and 4 for chief physician.
	Recommendation Heat (DReHeat)	Continuous variable, doctor's online recommendation popularity: 0–5 points.
	Hospital Level (HosLevel)	Categorical variable, representing the hospital tier where the doctor practices: 0 for unranked hospitals, 1 for Tier 1, 2 for Tier 1A, 3 for Tier 2, 4 for Tier 2A, 5 for Tier 3, and 6 for Tier 3A.
	Communication Time (CommuTime)	The total number of doctor–patient exchanges in one round of doctor–patient interaction.

4.3. Model Estimation

We employ Ordinary Least Squares (OLS) regression to estimate the data. The regression models are represented as follows:

The effect of linguistic features of patients' self-disclosure on doctors' informational support:

$$DInfoS_i = \beta_0 + \beta_1 Thorough_i + \beta_2 \text{Readability}_i + \beta_3 Professional_i + \beta_4 PSentiment_i + \beta_5 Control_i + \varepsilon_i \quad (1)$$

The effect of linguistic features of patients' self-disclosure on doctors' informational support:

$$DEmoS_i = \beta_0 + \beta_1 Thorough_i + \beta_2 \text{Readability}_i + \beta_3 Professional_i + \beta_4 PSentiment_i + \beta_5 Control_i + \varepsilon_i \quad (2)$$

5. Results

5.1. Descriptive Statistics and Correlation Analysis

We used Stata 17 to analyze the data. Tables 2 and 3 present the descriptive statistics and correlation matrix for the variables, respectively. Given the different scales of the variables, the independent and control variables were standardized, and a linear model was used to check the Variance Inflation Factor (VIF). As shown in Table 2, the VIF values are all less than 2, indicating that there was no significant multicollinearity among these variables.

Table 2. Descriptive statistics.

Variables	N	Mean	SD	Min	Max	VIF
DInfoS	52,730	88.12	96.29	0	2963	
DEmoS	52,730	0.765	1.206	0	25	
Completeness	52,730	354.4	321.4	2	7481	1.590
Readability	52,730	0.992	0.010	0.531	0.999	1.040
Expertise	52,730	0.075	0.030	0.002	0.500	1.070
PSentiment	52,730	−0.154	0.206	−1	1	1.070
PGender	52,730	0.637	0.481	0	1	1.030
PAge	52,730	49.12	14.79	1	90	1.100
DGender	52,730	0.808	0.394	0	1	1.070
DReHeat	52,730	3.916	0.533	1.700	5	1.480
DEmGrade	52,730	3.425	0.756	1	4	1.380
HosLevel	52,730	5.884	0.525	0	6	1.050
CommuTime	52,730	21.51	21.04	1	644	1.590

Table 3. Correlations of variables.

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13
1 DInfoS	1												
2 DEmoS	0.512 ***	1											
3 Completeness	0.294 ***	0.228 ***	1										
4 Readability	0.00200	0.011 ***	0.065 ***	1									
5 AHLiteracy	−0.098 ***	−0.120 ***	−0.110 ***	0.007 *	1								
6 PSentiment	−0.00600	−0.047 ***	−0.136 ***	−0.106 ***	−0.040 ***	1							
7 PGender	−0.046 ***	−0.010 **	−0.00200	−0.00300	0.069 ***	−0.008 **	1						
8 PAge	−0.099 ***	−0.083 ***	0.011 ***	0.024 ***	0.180 ***	−0.024 ***	0.108 ***	1					
9 DGender	−0.037 ***	0.074 ***	−0.040 ***	0.011 ***	0.035 ***	−0.039 ***	0.107 ***	0.118 ***	1				
10 DReHeat	0.015 ***	0.108 ***	0.084 ***	0.072 ***	0.090 ***	−0.063 ***	0.089 ***	0.216 ***	0.181 ***	1			
11 DEmGrade	0.00400	0.044 ***	0.103 ***	0.083 ***	0.061 ***	−0.060 ***	0.034 ***	0.125 ***	0.017 ***	0.501 ***	1		
12 HosLevel	−0.085 ***	−0.086 ***	−0.046 ***	0.024 ***	0.053 ***	0.00200	−0.014 ***	0.071 ***	−0.059 ***	0.169 ***	0.106 ***	1	
13 CommuTime	0.368 ***	0.319 ***	0.553 ***	−0.075 ***	−0.172 ***	0.086 ***	0.010 **	−0.031 ***	−0.015 ***	0.013 ***	−0.066 ***	−0.086 ***	1

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.2. Regression Results Analysis

Table 4 presents the ordinary least squares (OLS) regression results with robust standard errors. Model 1 examines the relationship between the linguistic features in patients' self-disclosure and doctors' provision of informational support (DInfoS). The results show that completeness (Completeness, $\beta = 11.063$, $p < 0.001$) and readability (Readability, $\beta = 1.544$, $p < 0.001$) have significant positive effects on doctors' informational support. This suggests that the more comprehensive and readable the patient's inquiry, the more informational support the doctor provides, supporting H1a and H2a. However, the coefficient of expertise (Expertise, $\beta = -1.055$, $p < 0.01$) is negative and statistically significant, suggesting that patients using more medical terminology and demonstrating higher medical expertise tend to receive less informational support from doctors. As a result, H3a is rejected. Additionally, positive sentiment (PSentiment, $\beta = -1.655$, $p < 0.001$) is negatively correlated with doctors' informational support, implying that the more positive a patient's sentiment, the less informational support they receive. Conversely, patients with more negative emotions receive more informational support, supporting H4a.

Table 4. Ordinary least squares (OLS) regression results.

Model Variables	Model 1 DInfoS	Model 2 DEmoS
Completeness	11.063 ***	0.047 ***
	(9.01)	(3.59)
Readability	1.544 ***	0.022 ***
	(3.64)	(4.40)
Expertise	−1.055 **	−0.074 ***
	(−2.71)	(−14.91)
PSentiment	−1.655 ***	−0.077 ***
	(−3.58)	(−13.98)
PGender	−8.205 ***	−0.043 ***
	(−9.70)	(−4.20)
PAge	−8.383 ***	−0.112 ***
	(−19.72)	(−22.01)
DGender	−6.159 ***	0.198 ***
	(−5.94)	(17.26)
DReHeat	2.685 ***	0.134 ***
	(6.08)	(21.20)
DEmGrade	1.382 **	0.025 ***
	(3.29)	(4.38)
HosLevel	−5.135 ***	−0.078 ***
	(−10.37)	(−11.72)
CommuTime	27.872 ***	0.342 ***
	(20.47)	(20.78)
Constant	98.319 ***	0.633 ***
	(94.99)	(55.66)
Observations	52,730	52,730
R-squared	0.155	0.139
F	215.9	281.9

Notes: Robust t-statistics in parentheses; ** $p < 0.01$, *** $p < 0.001$.

Model 2 investigates the effects of the linguistic features in patients' self-disclosure on doctors' provision of emotional support (DEmoS). The results reveal that completeness (Completeness, $\beta = 0.047$, $p < 0.001$) and readability (Readability, $\beta = 0.022$, $p < 0.001$) are positively associated with emotional support, indicating that patients who provide more comprehensive and easily understandable inquiries receive greater emotional support from doctors. Thus, H1b and H2b are supported. However, the coefficient of expertise (Expertise, $\beta = -0.074$, $p < 0.001$) is negatively significant, suggesting that patients displaying a higher level of medical expertise tend to receive less emotional support from doctors, leading to the rejection of H3b. Similarly, positive sentiment (PSentiment, $\beta = -0.077$, $p < 0.001$) is negatively correlated with emotional support, suggesting that the more positive a patient's sentiment, the less emotional support they receive, while patients with more negative emotions receive more emotional support, supporting H4b. Table 5 summarizes the hypothesis testing results.

Table 5. Summary of hypotheses testing.

No.	Hypotheses	Results
H1a	The completeness of patient self-disclosure positively influences doctors' informational support.	Supported
H1b	The completeness of patient self-disclosure positively influences doctors' emotional support.	Supported
H2a	The readability of patient self-disclosure positively influences doctors' informational support.	Supported
H2a	The readability of patient self-disclosure positively influences doctors' emotional support.	Supported
H3a	The expertise of patient self-disclosure positively influences doctors' informational support.	Not supported
H3b	The expertise of patient self-disclosure positively influences doctors' emotional support.	Not supported
H4a	The positive sentiment of patient self-disclosure negatively influences doctors' informational support.	Supported
H4b	The positive sentiment of patient self-disclosure negatively influences doctors' emotional support.	Supported

5.3. Robustness Checks

To ensure the robustness of our findings, we employed two different robustness testing methods, as summarized in Table 6.

Table 6. Robustness checks.

	Checks 1		Checks 2	
Model Variables	**Model 3 DInfoS**	**Model 4 DEmoS**	**Model 5 DTextLen**	**Model 6 DResNum**
Completeness	0.058 ***	0.030 ***	29.609 ***	0.436 ***
	(9.41)	(3.64)	(9.99)	(3.84)
Readability	0.018 ***	0.038 ***	3.496 ***	−0.088 ***
	(3.50)	(4.06)	(3.70)	(−3.64)
Expertise	−0.010 *	−0.115 ***	−2.678 **	−0.197 ***
	(−2.04)	(−14.83)	(−3.00)	(−9.10)
PSentiment	−0.023 ***	−0.097 ***	−2.395 *	0.185 ***
	(−4.87)	(−15.00)	(−2.28)	(6.08)
PGender	−0.098 ***	−0.056 ***	−20.091 ***	−0.194 ***
	(−11.03)	(−4.33)	(−10.31)	(−5.65)
PAge	−0.098 ***	−0.141 ***	−20.765 ***	−0.215 ***
	(−22.34)	(−21.71)	(−21.29)	(−12.40)
DGender	−0.049 ***	0.274 ***	−14.741 ***	−0.526 ***
	(−4.68)	(15.25)	(−6.10)	(−11.94)
DReHeat	0.043 ***	0.179 ***	8.034 ***	0.230 ***
	(8.50)	(24.01)	(7.69)	(10.45)
DEmGrade	0.031 ***	0.046 ***	0.849	−0.548 ***
	(6.87)	(5.91)	(0.85)	(−17.49)
HosLevel	−0.067 ***	−0.081 ***	−13.440 ***	−0.235 ***
	(−11.81)	(−13.52)	(−11.57)	(−8.59)
CommuTime	0.333 ***	0.340 ***	70.784 ***	5.758 ***
	(42.60)	(32.16)	(20.17)	(27.56)
lnalpha	−0.312 ***	−0.419 ***	-	-
	(−44.28)	(−18.06)		
Constant	4.508 ***	−0.578 ***	237.985 ***	9.564 ***
	(431.08)	(−32.27)	(99.16)	(221.02)
Observations	52,730	52,730	52,730	52,730
Log likelihood	−284,046	−60,380	-	-
R-squared	-	-	0.168	0.698
Pseudo R-squared	0.0165	0.0521	-	-
F	-	-	250.7	1719

Notes: Robust t-statistics in parentheses in Models 3 and 4; Robust z-statistics in parentheses in Models 5 and 6; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

First, given that both informational support (DInfoS) and emotional support (DEmoS) are count variables, we applied a negative binomial regression model instead of OLS to account for data overdispersion [79,80]. Model 3 presents the negative binomial regression results for informational support, while Model 4 reports the results for emotional support. The estimates remain consistent with those in Table 5, supporting the robustness of our findings.

Second, to verify the consistency of our results under alternative measures of doctors' service quality, we used text length per interaction and total number of responses as substitute indicators [14,81]. Model 5 reports the OLS regression results using text length as the dependent variable, which aligns with the main conclusions, reinforcing the robustness of our model. Model 6, which examines the effect of linguistic features on the number of responses, reveals a directional change in the coefficients for readability and

sentiment polarity. Specifically, higher readability of patient text is associated with fewer doctor responses, but this reduction in response frequency does not indicate a decrease in information provision. Instead, the length of doctors' replies increases, suggesting that when patient texts are more readable, doctors may respond less frequently but provide more detailed replies. On the other hand, more positive patient sentiment correlates with an increase in the number of doctor responses. One possible explanation is that doctors tend to be more cautious when addressing negative emotions, resulting in fewer but more extensive responses.

6. Discussion, Implications and Limitations

6.1. Key Findings

Drawing on signaling theory, this study examines the mechanisms through which linguistic features of patient self-disclosure influence doctors' service quality in an e-health context. The results reveal that completeness in patient self-disclosure has a significant positive effect on both informational and emotional support provided by doctors, aligning with previous research [23]. Additionally, readability in patient self-disclosure also has a positive effect on doctors' service quality, possibly because more detailed and clearly expressed information enhances doctors' understanding of patients' conditions and needs, enabling them to provide more precise support. Existing research has found that on short-term rental platforms, the readability of hosts' self-descriptions is positively associated with guests' perceived trust, thereby promoting booking behavior [28]. Building upon this, our study further explores the relationship between text readability and service quality in the context of doctor–patient interactions. In contrast, expertise in patient self-disclosure negatively impacts doctors' informational and emotional support, which contradicts our initial hypotheses. A possible explanation is that the excessive use of medical terminology reduces text readability, increasing the cognitive burden on doctors and ultimately affecting their service quality [82]. Moreover, patients' positive sentiment also negatively affects doctors' service quality. According to the negativity bias theory, doctors tend to allocate more attention and assistance to patients who exhibit negative emotions, whereas those displaying positive emotions may be perceived as requiring less intervention, leading to a reduction in doctors' service engagement. A prior study on offline clinical settings found that when patients express negative emotions or concerns, doctors tend to avoid engaging in emotional discussions, especially when patients explicitly express their concerns [32]. However, given the substantial differences between face-to-face and online consultations, doctors' responses to patient emotions may vary in digital settings. Effective face-to-face doctor–patient communication is shown to not only improve medical decision-making quality but also enhance patients' emotional well-being and other health-related outcomes [83]. Building on this foundation, this study extends the exploration of doctor–patient communication to the online consultation setting, examining how patients can adjust their communication strategies to enhance doctors' service quality and improve their overall consultation experience.

6.2. Implications

6.2.1. Theoretical Implications

This study offers several theoretical contributions.

First, by introducing signaling theory, this study explores how patients' linguistic signals influence doctors' online engagement in teleconsultations. The findings reveal that text completeness and readability facilitate higher-quality medical services, while expertise and positive sentiment negatively affect service quality. These results extend the application of signaling theory in doctor–patient interactions, demonstrating that patients'

linguistic expressions serve not only as a means of information transmission but also as crucial signals influencing doctors' responses. Furthermore, this study provides a new theoretical perspective on how optimizing patients' language expression can enhance healthcare service quality.

Second, existing research on online consultations has primarily focused on patient satisfaction and subsequent decision-making behaviors, with limited attention to patients' self-disclosure and linguistic features [16,64,84]. This study addresses this gap by examining completeness, readability, expertise, and sentiment polarity in patient self-disclosure from lexical, syntactic, semantic, and pragmatic dimensions, enriching the literature on patients' disclosure behavior and linguistic features.

Third, previous studies on doctor–patient interactions have mainly focused on how doctors' service quality influences patients, with limited exploration of the factors affecting doctors' service quality [13,85]. By focusing on patients' agency and proactivity, this study delves into doctor–patient dialogues to investigate how patients' linguistic expressions impact doctors' service quality, offering a novel perspective on online doctor–patient interactions.

Finally, employing real online consultation data, this study applies text mining and text classification techniques to accurately extract and quantify patients' linguistic features as well as doctors' informational and emotional support. This approach expands the application of natural language processing (NLP) in online healthcare research and provides a methodological reference for future studies on doctor–patient interactions.

6.2.2. Practical Implications

From a practical perspective, the findings of this study provide valuable insights for patients, doctors, and online consultation platforms in optimizing medical interactions.

First, this study reveals how linguistic features of patient self-disclosure influence doctors' service quality, offering insights into how patients can optimize their text expression in online consultations. The results indicate that a more comprehensive and readable description leads to greater social support from doctors. Therefore, patients should provide complete, well-structured descriptions of their medical conditions, avoiding missing or unclear information that may lead to misinterpretation or inefficient responses from doctors. Moreover, the excessive use of medical terms negatively impacts social support from doctors, suggesting that patients should use clear and accessible language to describe their symptoms. Additionally, emotional expression also influences doctors' responses—patients displaying highly positive emotions may receive less emotional support. Thus, patients should balance their emotional expression, appropriately conveying concerns and needs to ensure doctors fully understand their situation and provide adequate responses.

Second, the findings provide new insights for doctors, emphasizing the importance of recognizing variations in patients' linguistic expressions when delivering online medical services. Doctors should be more sensitive to patients' linguistic features, as differences in communication styles may unintentionally influence their judgment and responses. More importantly, they should prioritize patients' actual needs rather than relying solely on linguistic cues when determining the level of informational and emotional support required, ensuring medical fairness in online consultations.

Finally, the findings also provide valuable guidance for the optimization of online consultation platforms through technological and interface enhancements to facilitate more effective communication [86]. One possible improvement is the implementation of structured text input tools, which can help guide patients through step-by-step descriptions of their medical history, symptoms, onset time, and severity, allowing doctors to quickly grasp key information and improve consultation efficiency. Additionally, AI-assisted language optimization tools can be integrated to automatically refine patient input, such as

simplifying complex sentences, replacing obscure terms, and suggesting missing but essential information, thereby enhancing readability and completeness [87]. Another potential advancement involves emotion recognition technology, which could be used to analyze patients' emotional states and provide real-time indicators on the doctor's interface, encouraging them to offer more targeted emotional support when necessary. By incorporating these optimizations, online consultation platforms can enhance communication efficiency, improve user experience, and ultimately advance the quality of telemedicine services.

6.3. Limitations and Future Research Directions

This study has several limitations.

First, the analysis is based on data from a single online consultation platform in China and examines the impact of patients' self-disclosure on doctors' service quality in a Chinese-language context. However, linguistic and cultural differences may significantly influence self-disclosure and its effects on physician behavior. For example, Chinese and English differ significantly in language structure, information organization, and emotional expression, which could affect the patterns of self-disclosure and their impact on medical interactions. Consequently, our findings may primarily apply to Chinese-language consultations, and their generalizability to other linguistic contexts remains uncertain. Future research could replicate this study using data from multiple countries and platforms to enhance the robustness of the conclusions. Additionally, cross-cultural and cross-linguistic comparisons could provide valuable insights into how language and cultural backgrounds shape patient communication and doctor–patient interactions, offering a broader perspective for global online healthcare research.

Second, this study focuses exclusively on text-based online consultations, without considering telephone or other telemedicine modalities. In settings involving voice communication, paralinguistic cues such as tone and speech rate may impact doctors' responses and engagement. Future studies could expand to audio and video-based consultations to investigate the influence of vocal and visual signals in doctor–patient interactions, contributing to a more comprehensive understanding of communication mechanisms in online healthcare settings.

Third, this study has methodological limitations related to variable measurement. Doctors' service quality is evaluated based on informational and emotional support. However, due to dataset limitations, we can not incorporate indicators of how well patient needs were met, such as patient ratings and satisfaction scores. Additionally, potential biases in text data analysis should be acknowledged, as variable extraction may be influenced by interpretation bias or algorithmic constraints. Future research could employ more refined measurement approaches to improve the accuracy of variable assessment, thereby enhancing the reliability and external validity of the findings.

Finally, although this study focuses on gastritis and liver cancer based on disease severity, different diseases may involve varying patient information needs and communication styles. Future research could incorporate a wider range of diseases to strengthen the generalizability of the findings and further explore how patient linguistic features influence doctors' service quality across different medical conditions.

Author Contributions: Conceptualization, M.P. and X.Y.; methodology, M.P. and K.Z.; formal analysis, K.Z., Y.G. and K.S.; data curation, K.S.; writing—original draft preparation, M.P.; writing—review and editing, D.G. and X.Y.; funding acquisition, D.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 72131006, 72271082, 72071063) and the Natural Science Foundation of Anhui Province (Grant No. 2408085J041).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable. We would like to clarify that our study is based on publicly available online consultation data obtained from a widely used Chinese online healthcare platform, with all data processing conducted in compliance with the platform's terms of service and relevant ethical guidelines. Therefore, informed consent was not required.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1. Text Classification Process for Sentiment Polarity and Emotional Support

The text classification process consists of four main steps: defining annotation rules and sample labeling, text preprocessing, feature extraction and vectorization, and classifier construction and evaluation. The specific workflow is as follows:

1. Defining Annotation Rules and Sample Labeling

We first established annotation rules for identifying patients' sentiment polarity and doctors' emotional support based on unstructured text data from doctor–patient interactions. Sentimet polarity was categorized into negative, neutral, and positive, while doctors' emotional support was characterized by expressions of reassurance, encouragement, or empathy, such as "Don't worry", "I understand you", and "Stay strong". Specific annotation rules are shown in Table A1.

We recruited three medical graduate students for the annotation process: First, the annotators compiled an initial lexicon based on the specified annotation rules and the context of the consultations. Then, each student randomly selected 1000 patient texts and 1000 doctor texts for annotation. Cross-validation was conducted to assess annotation consistency, and any disagreements were discussed to refine the rules. Finally, after finalizing the annotation rules, we randomly selected 3000 doctor texts and 3000 patient texts for formal annotation. Each annotator labeled 2000 texts, and to ensure reliability, all annotations were cross-checked so that each text was labeled by two annotators.

2. Text Preprocessing

We used the jieba package in Python to segment the text data and removed stop words based on a stop word dictionary, resulting in segmented text.

3. Feature Extraction and Vectorization

To represent textual features effectively, different vectorization methods were applied based on the classification model. (a). Support Vector Machine (SVM), used Term Frequency-Inverse Document Frequency (TF-IDF) to compute word weights for text representation. (b). Text Convolutional Neural Network (TextCNN), trained a SkipGram word vector model and utilized the trained Word2Vec model for word embedding, converting natural language into word vectors. (c). Bidirectional Encoder Representations from Transformers (BERT), loaded the pre-trained BERT model, which directly transformed text into word vector through its embedding layer.

4. Classifier Construction and Evaluation

The labeled dataset was split into 80% training data and 20% testing data. We implemented three classification models. We used three models for text classification: SVM, TextCNN, and BERT. The models were trained using the training set and evaluated on the test set. As shown in Table A2, BERT outperformed SVM and TextCNN across all key

performance metrics, including accuracy, precision, recall, and F1-score. Therefore, BERT was selected as the final classifier for automatically labeling the unlabeled dataset.

Table A1. The labeling rules of sentiment polarity and emotional support.

Variables	Classification	Keywords	Example	Tags
Patient's Sentiment Polarity	Positive	Positive words like happy, peaceful, grateful, love, etc.	"I have great faith in the future."	1
	Neutral		"Which kinds of medicine to take?"	0
	Negative	Negative words like worry, sad, depressed, etc.	"I'm afraid, please help me."	−1
Doctor's Emotional Support Provision	Yes	don't give up, understand, sympathize, come on, be optimistic, etc.	"Let's face it together"	1
	No		"Well, keep the fluids flowing and give your intestines a rest"	0

Table A2. Classifier performance evaluation results.

Classifiers	Sentiment Polarity				Emotional Support			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
SVM	0.81	0.78	0.38	0.39	0.87	0.89	0.84	0.86
TextCNN	0.82	0.61	0.43	0.46	0.78	0.36	0.46	0.40
BERT	0.90	0.82	0.74	0.77	0.96	0.95	0.95	0.94

References

1. Yang, X.; Xi, N.; Gu, D.; Liang, C.; Liu, H.; Tang, H.; Hamari, J. Medical Practice in Gamified Online Communities: Longitudinal Effects of Gamification on Doctor Engagement. *Inf. Manag.* **2023**, *61*, 103906. [CrossRef]
2. Fan, W.; Jiang, Y.; Pei, J.; Yan, P.; Qiu, L. The Impact of Medical Insurance Payment Systems on Patient Choice, Provider Behavior, and Out-of-pocket Rate: Fee-for-service versus Diagnosis-related Groups. *Decis. Sci.* **2024**, *55*, 245–261. [CrossRef]
3. Liu, Y.; Kong, Q.; Wang, S.; Zhong, L.; van de Klundert, J. The Impact of Hospital Attributes on Patient Choice for First Visit: Evidence from a Discrete Choice Experiment in Shanghai, China. *Health Policy Plan.* **2020**, *35*, 267–278.
4. Li, J.; Tang, J.; Jiang, L.; Yen, D.C.; Liu, X. Economic Success of Physicians in the Online Consultation Market: A Signaling Theory Perspective. *Int. J. Electron. Commer.* **2019**, *23*, 244–271. [CrossRef]
5. Wu, Q.; Xie, X.; Liu, W.; Wu, Y. Implementation Efficiency of the Hierarchical Diagnosis and Treatment System in China: A Case Study of Primary Medical and Health Institutions in Fujian Province. *Int. J. Health Plann. Manag.* **2022**, *37*, 214–227.
6. Zhang, W.; Zhou, F.; Fei, Y. Repetitions in Online Doctor-Patient Communication: Frequency, Functions, and Reasons. *Patient Educ. Couns.* **2023**, *107*, 107565.
7. Shreve, E.G.; Harrigan, J.A.; Kues, J.R.; Kagas, D.K. Nonverbal Expressions of Anxiety in Physician-Patient Interactions. *Psychiatry* **1988**, *51*, 378–384. [CrossRef]
8. Tan, H.; Yan, M. Physician-User Interaction and Users' Perceived Service Quality: Evidence from Chinese Mobile Healthcare Consultation. *Inf. Technol. People* **2020**, *33*, 1403–1426. [CrossRef]
9. Gu, D.; Li, T.; Wang, X.; Yang, X.; Yu, Z. Visualizing the Intellectual Structure and Evolution of Electronic Health and Telemedicine Research. *Int. J. Med. Inform.* **2019**, *130*, 103947. [CrossRef]
10. Dagger, T.S.; Sweeney, J.C.; Johnson, L.W. A Hierarchical Model of Health Service Quality: Scale Development and Investigation of an Integrated Model. *J. Serv. Res.* **2007**, *10*, 123–142.
11. Akter, S.; D'ambra, J.; Ray, P. Development and Validation of an Instrument to Measure User Perceived Service Quality of mHealth. *Inf. Manag.* **2013**, *50*, 181–195. [CrossRef]
12. Yang, Y.; Zhang, X.; Lee, P.K. Improving the Effectiveness of Online Healthcare Platforms: An Empirical Study with Multi-Period Patient-Doctor Consultation Data. *Int. J. Prod. Econ.* **2019**, *207*, 70–80. [CrossRef]
13. Yang, H.; Guo, X.; Wu, T. Exploring the Influence of the Online Physician Service Delivery Process on Patient Satisfaction. *Decis. Support Syst.* **2015**, *78*, 113–121. [CrossRef]
14. Wang, Q.; Qiu, L.; Xu, W. Informal Payments and Doctor Engagement in an Online Health Community: An Empirical Investigation Using Generalized Synthetic Control. *Inf. Syst. Res.* **2023**, *35*, 706–726. [CrossRef]
15. Lu, N.; Wu, H. Exploring the Impact of Word-of-Mouth about Physicians' Service Quality on Patient Choice Based on Online Health Communities. *BMC Med. Inform. Decis. Mak.* **2016**, *16*, 151. [CrossRef]

16. Chen, S.; Guo, X.; Wu, T.; Ju, X. Exploring the Online Doctor-Patient Interaction on Patient Satisfaction Based on Text Mining and Empirical Analysis. *Inf. Process. Manag.* **2020**, *57*, 102253. [CrossRef]
17. He, Y.; Guo, X.; Wu, T.; Vogel, D. The Effect of Interactive Factors on Online Health Consultation Review Deviation: An Empirical Investigation. *Int. J. Med. Inform.* **2022**, *163*, 104781. [CrossRef]
18. Cao, X.; Liu, Y.; Zhu, Z.; Hu, J.; Chen, X. Online Selection of a Physician by Patients: Empirical Study from Elaboration Likelihood Perspective. *Comput. Hum. Behav.* **2017**, *73*, 403–412. [CrossRef]
19. Zhang, X.; Guo, F.; Xu, T.; Li, Y. What Motivates Physicians to Share Free Health Information on Online Health Platforms? *Inf. Process. Manag.* **2020**, *57*, 102166. [CrossRef]
20. Yang, H.; Du, H.S.; He, W.; Qiao, H. Understanding the Motivators Affecting Doctors' Contributions in Online Healthcare Communities: Professional Status as a Moderator. *Behav. Inf. Technol.* **2021**, *40*, 146–160. [CrossRef]
21. Guo, S.; Guo, X.; Fang, Y.; Vogel, D. How Doctors Gain Social and Economic Returns in Online Health-Care Communities: A Professional Capital Perspective. *J. Manag. Inf. Syst.* **2017**, *34*, 487–519. [CrossRef]
22. Wang, Y.; Wu, H.; Xia, C.; Lu, N. Impact of the Price of Gifts From Patients on Physicians' Service Quality in Online Consultations: Empirical Study Based on Social Exchange Theory. *J. Med. Internet Res.* **2020**, *22*, e15685. [CrossRef]
23. Liu, J.; He, J.; He, S.; Li, C.; Yu, C.; Li, Q. Patients' Self-Disclosure Positively Influences the Establishment of Patients' Trust in Physicians: An Empirical Study of Computer-Mediated Communication in an Online Health Community. *Front. Public Health* **2022**, *10*, 823692.
24. Chiu, C.-M.; Hsu, M.-H.; Wang, E.T.G. Understanding Knowledge Sharing in Virtual Communities: An Integration of Social Capital and Social Cognitive Theories. *Decis. Support Syst.* **2006**, *42*, 1872–1888.
25. Parhankangas, A.; Renko, M. Linguistic Style and Crowdfunding Success among Social and Commercial Entrepreneurs. *J. Bus. Ventur.* **2017**, *32*, 215–236.
26. Johnson, S.L.; Safadi, H.; Faraj, S. The Emergence of Online Community Leadership. *Inf. Syst. Res.* **2015**, *26*, 165–187. [CrossRef]
27. Hartzler, A.; Pratt, W. Managing the Personal Side of Health: How Patient Expertise Differs from the Expertise of Clinicians. *J. Med. Internet Res.* **2011**, *13*, e1728. [CrossRef]
28. Zhang, L.; Yan, Q.; Zhang, L. A Text Analytics Framework for Understanding the Relationships among Host Self-Description, Trust Perception and Purchase Behavior on Airbnb. *Decis. Support Syst.* **2020**, *133*, 113288. [CrossRef]
29. Zhou, J.; Zhang, Q.; Zhou, S.; Li, X.; Michael Zhang, X. Unintended Emotional Effects of Online Health Communities: A Text Mining-Supported Empirical Study. *MIS Q.* **2023**, *47*, 195–226. [CrossRef]
30. Chen, L.; Baird, A.; Straub, D. A Linguistic Signaling Model of Social Support Exchange in Online Health Communities. *Decis. Support Syst.* **2020**, *130*, 113233. [CrossRef]
31. Al Salman, A.; Kim, A.; Mercado, A.; Ring, D.; Doornberg, J.; Fatehi, A.; Crijns, T.J. Are Patient Linguistic Tones Associated with Mental Health and Perceived Clinician Empathy? *J. Bone Jt. Surg.* **2021**, *103*, 2181. [CrossRef]
32. Mjaaland, T.A.; Finset, A.; Jensen, B.F.; Gulbrandsen, P. Physicians' Responses to Patients' Expressions of Negative Emotions in Hospital Consultations: A Video-Based Observational Study. *Patient Educ. Couns.* **2011**, *84*, 332–337. [CrossRef]
33. Spence, M. Job Market Signaling. *Q. J. Econ.* **1973**, *87*, 355–374. [CrossRef]
34. Cheung, C.M.K.; Xiao, B.S.; Liu, I.L.B. Do Actions Speak Louder than Voices? The Signaling Role of Social Information Cues in Influencing Consumer Purchase Decisions. *Decis. Support Syst.* **2014**, *65*, 50–58. [CrossRef]
35. Wells, J.D.; Valacich, J.S.; Hess, T.J. What Signal Are You Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions. *MIS Q.* **2011**, *35*, 373. [CrossRef]
36. Choi, H.S.; Ko, M.S.; Medlin, D.; Chen, C. The Effect of Intrinsic and Extrinsic Quality Cues of Digital Video Games on Sales: An Empirical Investigation. *Decis. Support Syst.* **2018**, *106*, 86–96. [CrossRef]
37. Benlian, A.; Hess, T. The Signaling Role of IT Features in Influencing Trust and Participation in Online Communities. *Int. J. Electron. Commer.* **2011**, *15*, 7–56. [CrossRef]
38. Kiyohara, L.Y.; Kayano, L.K.; Kobayashi, M.L.T.; Alessi, M.S.; Yamamoto, M.U.; Yunes-Filho, P.R.M.; Pessoa, R.R.; Mandelbaum, R.; Okubo, S.T.; Watanuki, T.; et al. The Patient-Physician Interactions as Seen by Undergraduate Medical Students. *Sao Paulo Med. J.* **2001**, *119*, 97–100. [CrossRef]
39. Connelly, B.L.; Certo, S.T.; Ireland, R.D.; Reutzel, C.R. Signaling Theory: A Review and Assessment. *J. Manag.* **2011**, *37*, 39–67. [CrossRef]
40. Jiang, S.; Liu, X.; Chi, X. Effect of Writing Style on Social Support in Online Health Communities: A Theoretical Linguistic Analysis Framework. *Inf. Manag.* **2022**, *59*, 103683. [CrossRef]
41. Khurana, S.; Qiu, L.; Kumar, S. When a Doctor Knows, It Shows: An Empirical Analysis of Doctors' Responses in a Q&A Forum of an Online Healthcare Portal. *Inf. Syst. Res.* **2019**, *30*, 872–891. [CrossRef]
42. Yang, H.; Guo, X.; Wu, T.; Ju, X. Exploring the Effects of Patient-Generated and System-Generated Information on Patients' Online Search, Evaluation and Decision. *Electron. Commer. Res. Appl.* **2015**, *14*, 192–203. [CrossRef]

43. Ma, J.; Gao, S.; Wang, P.; Liu, Y. High Level of Self-Disclosure on SNSs Facilitates Cooperation: A Serial Mediation Model of Psychological Distance and Trust. *Comput. Hum. Behav.* **2024**, *150*, 107976. [CrossRef]
44. Finkel, E.J.; Norton, M.I.; Reis, H.T.; Ariely, D.; Caprariello, P.A.; Eastwick, P.W.; Frost, J.H.; Maniaci, M.R. When Does Familiarity Promote Versus Undermine Interpersonal Attraction? A Proposed Integrative Model From Erstwhile Adversaries. *Perspect. Psychol. Sci.* **2015**, *10*, 3–19. [CrossRef] [PubMed]
45. Utz, S. The Function of Self-Disclosure on Social Network Sites: Not Only Intimate, but Also Positive and Entertaining Self-Disclosures Increase the Feeling of Connection. *Comput. Hum. Behav.* **2015**, *45*, 1–10. [CrossRef]
46. El Ouirdi, M.; Segers, J.; El Ouirdi, A.; Pais, I. Predictors of Job Seekers' Self-Disclosure on Social Media. *Comput. Hum. Behav.* **2015**, *53*, 1–12. [CrossRef]
47. Bae, S.; Jang, J.; Kim, J. Good Samaritans on Social Network Services: Effects of Shared Context Information on Social Supports for Strangers. *Int. J. Hum.-Comput. Stud.* **2013**, *71*, 900–918. [CrossRef]
48. Krasnova, H.; Veltri, N.F.; Günther, O. Self-Disclosure and Privacy Calculus on Social Networking Sites: The Role of Culture. *Bus. Inf. Syst. Eng.* **2012**, *4*, 127–135. [CrossRef]
49. Villarroel Ordenes, F.; Grewal, D.; Ludwig, S.; Ruyter, K.D.; Mahr, D.; Wetzels, M. Cutting through Content Clutter: How Speech and Image Acts Drive Consumer Sharing of Social Media Brand Messages. *J. Consum. Res.* **2019**, *45*, 988–1012. [CrossRef]
50. Ansari, S.; Gupta, S. Customer Perception of the Deceptiveness of Online Product Reviews: A Speech Act Theory Perspective. *Int. J. Inf. Manag.* **2021**, *57*, 102286. [CrossRef]
51. Tweedie, F.J.; Baayen, R.H. How Variable May a Constant Be? Measures of Lexical Richness in Perspective. *Comput. Humanit.* **1998**, *32*, 323–352. [CrossRef]
52. Castro, C.M.; Wilson, C.; Wang, F.; Schillinger, D. Babel Babble: Physicians' Use of Unclarified Medical Jargon with Patients. *Am. J. Health Behav.* **2007**, *31*, S85–S95. [PubMed]
53. Yin, D.; Bond, S.D.; Zhang, H. Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews. *MIS Q.* **2014**, *38*, 539–560. [CrossRef]
54. Ludwig, S.; de Ruyter, K.; Mahr, D.; Wetzels, M.; Brüggen, E.; de Ruyck, T. Take Their Word for It: The Symbolic Role of Linguistic Style Matches in User Communities. *MIS Q.* **2014**, *38*, 1201–1218.
55. Ouyang, P.; Wang, J.-J.; Jasmine Chang, A.-C. Patients Need Emotional Support: Managing Physician Disclosure Information to Attract More Patients. *Int. J. Med. Inform.* **2022**, *158*, 104674. [CrossRef]
56. Kasimtseva, L.; Kiseleva, L.; Dzhabrailova, S. Doctor-Patient Communication as a Linguistic Model. In Proceedings of the 1st International Scientific Practical Conference "The Individual and Society in the Modern Geopolitical Environment" (ISMGE 2019), Volgograd, Russian, 23–29 May 2019; Atlantis Press: Dordrecht, The Netherlands, 2019. [CrossRef]
57. Bientzle, M.; Griewatz, J.; Kimmerle, J.; Küppers, J.; Cress, U.; Lammerding-Koeppel, M. Impact of Scientific Versus Emotional Wording of Patient Questions on Doctor-Patient Communication in an Internet Forum: A Randomized Controlled Experiment with Medical Students. *J. Med. Internet Res.* **2015**, *17*, e268. [CrossRef]
58. Greaves, F.; Ramirez-Cano, D.; Millett, C.; Darzi, A.; Donaldson, L. Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online. *J. Med. Internet Res.* **2013**, *15*, e239. [CrossRef]
59. Gu, D.; Li, M.; Yang, X.; Gu, Y.; Zhao, Y.; Liang, C.; Liu, H. An Analysis of Cognitive Change in Online Mental Health Communities: A Textual Data Analysis Based on Post Replies of Support Seekers. *Inf. Process. Manag.* **2023**, *60*, 103192. [CrossRef]
60. Wang, X.; High, A.; Wang, X.; Zhao, K. Predicting Users' Continued Engagement in Online Health Communities from the Quantity and Quality of Received Support. *J. Assoc. Inf. Sci. Technol.* **2021**, *72*, 710–722. [CrossRef]
61. Liu, X.; Chi, X.; Li, J.; Zhou, S.; Cheng, Y. Doctors' Self-Presentation Strategies and the Effects on Patient Selection in Psychiatric Department from an Online Medical Platform: A Combined Perspective of Impression Management and Information Integration. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 13. [CrossRef]
62. Yao, T.; Zheng, Q.; Fan, X. The Impact of Online Social Support on Patients' Quality of Life and the Moderating Role of Social Exclusion. *J. Serv. Res.* **2015**, *18*, 369–383. [CrossRef]
63. Liu, S.; Si, G.; Gao, B. Which Voice Are You Satisfied with? Understanding the Physician-Patient Voice Interactions on Online Health Platforms. *Decis. Support Syst.* **2022**, *157*, 113754. [CrossRef]
64. Zhang, X.; Guo, X.; Lai, K.; Yi, W. How Does Online Interactional Unfairness Matter for Patient-Doctor Relationship Quality in Online Health Consultation? The Contingencies of Professional Seniority and Disease Severity. *Eur. J. Inf. Syst.* **2019**, *28*, 336–354. [CrossRef]
65. Park, S.; Chung, N. Mediating Roles of Self-Presentation Desire in Online Game Community Commitment and Trust Behavior of Massive Multiplayer Online Role-Playing Games. *Comput. Hum. Behav.* **2011**, *27*, 2372–2379. [CrossRef]
66. Lu, X.; Jiang, J.; Head, M.; Yang, J. The Impact of Linguistic Complexity on Leadership in Online Q&A Communities: Comparing Knowledge Shaping and Knowledge Adding. *Inf. Manag.* **2022**, *59*, 103675. [CrossRef]
67. Bernhardt, J.M.; Brownfield, E.D.; Parker, R.M.; Schwartzberg, J.G.; VanGeest, J.B.; Wang, C.C. Understanding Health Literacy: Implications for Medicine and Public Health. *Chic. Am. Med. Assoc.* **2005**, *98*, 980–981.

68. Mackert, M.; Mabry-Flynn, A.; Champlin, S.; Donovan, E.E.; Pounders, K. Health Literacy and Health Information Technology Adoption: The Potential for a New Digital Divide. *J. Med. Internet Res.* **2016**, *18*, e6349. [CrossRef] [PubMed]
69. Howard, D.J.; Gengler, C. Emotional Contagion Effects on Product Attitudes: Figure 1. *J. Consum. Res.* **2001**, *28*, 189–201. [CrossRef]
70. Park, C.S. Applying "Negativity Bias" to Twitter: Negative News on Twitter, Emotions, and Political Learning. *J. Inf. Technol. Polit.* **2015**, *12*, 342–359. [CrossRef]
71. Naing, S.Z.S.; Udomwong, P. Public Opinions on ChatGPT: An Analysis of Reddit Discussions by Using Sentiment Analysis, Topic Modeling, and SWOT Analysis. *Data Intell.* **2024**, *6*, 344–374. [CrossRef]
72. Eslami, S.P.; Ghasemaghaei, M.; Hassanein, K. Which Online Reviews Do Consumers Find Most Helpful? A Multi-Method Investigation. *Decis. Support Syst.* **2018**, *113*, 32–42. [CrossRef]
73. Wu, H.; Lu, N. Online Written Consultation, Telephone Consultation and Offline Appointment: An Examination of the Channel Effect in Online Health Communities. *Int. J. Med. Inform.* **2017**, *107*, 107–119. [CrossRef]
74. Liu, Q.; Wang, B.; Kong, Y.; Cheng, K.K. China's Primary Health-Care Reform. *Lancet* **2011**, *377*, 2064–2066. [CrossRef]
75. Hu, Y.; Zhang, Z. Skilled Doctors in Tertiary Hospitals Are Already Overworked in China. *Lancet Glob. Health* **2015**, *3*, e737. [CrossRef] [PubMed]
76. Fan, W.; Zhou, Q.; Qiu, L.; Kumar, S. Should Doctors Open Online Consultation Services? An Empirical Investigation of Their Impact on Offline Appointments. *Inf. Syst. Res.* **2023**, *34*, 629–651. [CrossRef]
77. Sawicki, J.; Ganzha, M.; Paprzycki, M. The State of the Art of Natural Language Processing—A Systematic Automated Review of NLP Literature Using NLP Techniques. *Data Intell.* **2023**, *5*, 707–749. [CrossRef]
78. Deng, Z.; Deng, Z.; Fan, G.; Wang, B.; Fan, W.; Liu, S. More Is Better? Understanding the Effects of Online Interactions on Patients Health Anxiety. *J. Assoc. Inf. Sci. Technol.* **2023**, *74*, 1243–1264. [CrossRef]
79. Akkas, A.; Gaur, V.; Simchi-Levi, D. Drivers of Product Expiration in Consumer Packaged Goods Retailing. *Manag. Sci.* **2019**, *65*, 2179–2195. [CrossRef]
80. Li, Y.; Zhao, L. Collaborating with Bounty Hunters: How to Encourage White Hat Hackers' Participation in Vulnerability Crowdsourcing Programs through Formal and Relational Governance. *Inf. Manag.* **2022**, *59*, 103648. [CrossRef]
81. Jing, D.; Jin, Y.; Liu, J. The Impact of Monetary Incentives on Physician Prosocial Behavior in Online Medical Consulting Platforms: Evidence From China. *J. Med. Internet Res.* **2019**, *21*, e14685. [CrossRef]
82. Liu, X.; Wang, G.A.; Fan, W.; Zhang, Z. Finding Useful Solutions in Online Knowledge Communities: A Theory-Driven Design and Multilevel Analysis. *Inf. Syst. Res.* **2020**, *31*, 731–752. [CrossRef]
83. Street, R.L.; Makoul, G.; Arora, N.K.; Epstein, R.M. How Does Communication Heal? Pathways Linking Clinician–Patient Communication to Health Outcomes. *Patient Educ. Couns.* **2009**, *74*, 295–301. [CrossRef] [PubMed]
84. Liu, S.; Wang, H.; Gao, B.; Deng, Z. Doctors' Provision of Online Health Consultation Service and Patient Review Valence: Evidence from a Quasi-Experiment. *Inf. Manag.* **2022**, *59*, 103360. [CrossRef]
85. Liu, S.; Zhang, M.; Gao, B.; Jiang, G. Physician Voice Characteristics and Patient Satisfaction in Online Health Consultation. *Inf. Manag.* **2020**, *57*, 103233. [CrossRef]
86. Liang, C.; Wang, X.; Gu, D.; Li, P.; Chen, H.; Xu, Z. Smart Management Information Systems (Smis): Concept, Evolution, Research Hotspots and Applications. *Data Intell.* **2023**, *5*, 857–884. [CrossRef]
87. Zhao, W.; Gu, D.; Yang, X.; Jia, M.; Liang, C.; Wang, X.; Zolotarev, O. MedT2T: An Adaptive Pointer Constrain Generating Method for a New Medical Text-to-Table Task. *Future Gener. Comput. Syst.* **2024**, *161*, 586–600. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

systems

MDPI

Article

Big-Data-Assisted Urban Governance: A Machine-Learning-Based Data Record Standard Scoring Method

Zicheng Zhang [1,*] and Tianshu Zhang [2]

[1] School of Modern Post, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
[2] School of Information Management, Nanjing University, Nanjing 210023, China; zhangtianshu@hhu.edu.cn
* Correspondence: 20230173@njupt.edu.cn

Abstract: With the increasing adoption of digital governance and big data analytics, the quality of government hotline data significantly affects urban governance and public service efficiency. However, existing methods for assessing data record standards focus predominantly on structured data, exhibiting notable inadequacies in handling the complexities inherent in unstructured or semi-structured textual hotline records. To address these shortcomings, this study develops a comprehensive scoring method tailored for evaluating multi-dimensional data record standards in government hotline data. By integrating advanced deep learning models, we systematically analyze six evaluation indicators: classification predictability, dispatch accuracy, record correctness, address accuracy, adjacent sentence similarity, and full-text similarity. Empirical analysis reveals a significant positive correlation between improved data record standards and higher work order completion rates, particularly highlighting the crucial role of semantic-related indicators (classification predictability and adjacent sentence similarity). Furthermore, the results indicate that the work order field strengthens the positive impact of data standards on completion rates, whereas variations in departmental data-handling capabilities weaken this relationship. This study addresses existing inadequacies by proposing a novel scoring method emphasizing semantic measures and provides practical recommendations—including standardized language usage, intelligent analytic support, and targeted staff training—to effectively enhance urban governance.

Keywords: data record standard; government hotline; text classification; BERT; data quality

1. Introduction

With the rapid advancement of information technology, enhancing national modern governance capabilities has become a top priority. Establishing robust communication channels between the government and the public is fundamental to achieving this goal [1]. Government hotlines, through both telephone and internet channels, serve as crucial platforms for collecting public complaints, suggestions, and feedback on public affairs. These hotlines help address public concerns in a timely manner, improve administrative transparency, and enhance the quality of public services [2].

In China, the State Council's guidelines on optimizing public service mechanisms explicitly call for strengthening public feedback systems and improving the responsiveness and accuracy of government services through data-driven means [3]. Furthermore, government hotline data have been officially recognized as a core part of digital government construction, with the National Development and Reform Commission and the Ministry of Industry and Information Technology issuing documents that emphasize the importance of high-quality data management in public service systems [4].

A recent survey shows that 63% of cities have established clear development goals for their government hotline departments, aiming to create unified and accessible service platforms. However, these strategies primarily emphasize departmental growth while lacking targeted plans for data governance. This gap underscores the pressing need to enhance the quality of hotline data to ensure accurate issue handling and improve the effectiveness of public service delivery [5]. More broadly, the quality of government data plays a pivotal role in enabling effective social governance through big data. Without sufficient quality, even large volumes of data become difficult to utilize or analyze, providing limited value for policy-making and public administration. Therefore, evaluating and improving data quality remains a core challenge for governments and a critical factor in the success of digital governance systems [6].

In recent years, scholars have conducted extensive research on government hotlines. Wang et al. [7] proposed a classification model to optimize the routing of hotline issues to relevant departments. Li et al. [8] developed a multi-view deep learning framework to improve dispatch accuracy, while Zhang et al. [9] used text features to predict dispatch results. Other studies have explored multi-label classification methods [10], topic analysis of public concerns [11], and citizen satisfaction during crisis periods such as the COVID-19 pandemic [12].

However, current research predominantly focuses on improving the accuracy of issue classification and dispatch, with limited attention to the quality assessment of hotline records themselves. As Zhang et al. [9] pointed out, the effectiveness of classification and dispatch is heavily dependent on the completeness, consistency, and clarity of the original data input. Karkošková [13] emphasized that data quality assessments are foundational for ensuring reliability in policy analysis and service design, particularly in complex and multi-sectoral governance systems.

Commercial call centers often employ standardized tools such as Service-Level Agreements (SLAs) and Customer Satisfaction Indices (CSIs) to evaluate service effectiveness [14,15]. However, these frameworks focus on response speed or satisfaction metrics and are not well suited to evaluating the semantic and structural quality of records in the public sector. Government hotline data involve sensitive, multi-departmental issues that must meet regulatory, compliance, and ethical standards [16].

Moreover, as shown in Figure 1, the typical government hotline workflow—from call reception and information entry to issue dispatch and resolution—demands strict adherence to data recording protocols. Any deviation from these standards can disrupt task assignment and reduce system effectiveness. Ensuring standardization and consistency in data entry is not only essential for accurate issue handling but also for maintaining public trust in government responsiveness.

Government hotline data are characterized by their unstructured nature, containing rich semantic information that complicates traditional data governance methods. Unlike structured public data, such as census data [17], hotline records are composed of text that includes complaints, suggestions, and feedback, often involving multiple domains like public administration, legal issues, and social governance [18]. This makes the data complex and challenging to process, as they contain informal, ambiguous, and emotionally charged language, which cannot be easily quantified using standard techniques. Given the complexity and richness of the data, traditional quantitative methods are insufficient for capturing the nuances inherent in these textual records. To address this, a more advanced, text-based quantification method is necessary—one that not only processes the data but also captures their underlying semantic meaning. This approach stands in contrast to simpler forms of public data governance, where data are typically more structured and easier to quantify. The necessity for a specialized method to quantify and evaluate the

quality of government hotline data—particularly using a semantic-based approach—forms the core contribution of this study.

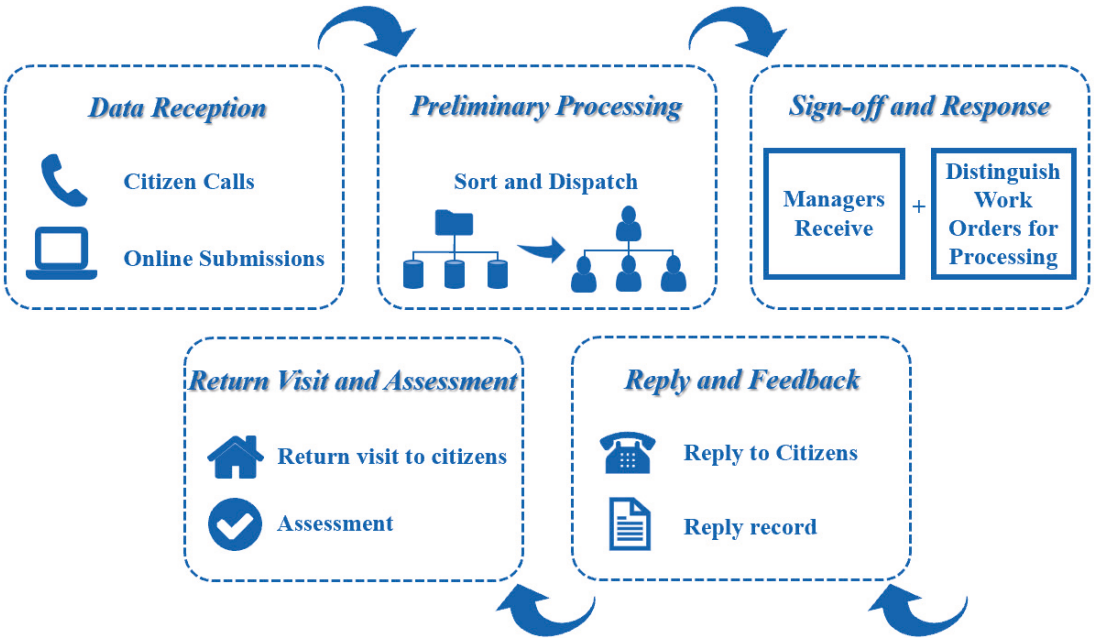


Figure 1. Government hotline processing workflow.

The objective of this study is to develop and implement a standardized scoring system for evaluating the quality of government hotline data, addressing the shortcomings in current research, where data quality assessment has been largely overlooked. Low-quality data records can lead to biased decision-making and negatively impact service quality [19]. Specifically, this study focuses on the methodology for automating the evaluation of hotline data quality using a customized scoring system. This system is designed to assess the completeness, consistency, and accuracy of hotline records, ensuring that they adhere to established standards. By implementing this system, this paper aims to enhance the effectiveness of government decision-making, improve public service delivery, and strengthen public trust in government responsiveness.

2. Literature Review

2.1. Current Research Status

Existing methods for the assessment of data quality typically focus on several key dimensions, including completeness, consistency, accuracy, timeliness, and reliability of data [20,21]. These dimensions are foundational in evaluating data quality across various fields, but there is a need for a more systematic approach to categorizing and assessing data quality frameworks. Schmidt et al. [22] proposed a data quality assessment framework specifically for the health research field. This framework emphasizes completeness, consistency, and accuracy as the core dimensions for ensuring the reliability and validity of medical data. It adopts a multi-level, multi-dimensional analysis that enables thorough evaluation across different layers of data.

In the industrial domain, Song et al. [23] introduced the Industrial Multivariate Time Series Data Quality Assessment (IMTSDQA) method, which builds upon traditional data quality evaluations by introducing fine-grained sub-dimensions within key categories such as completeness, normativeness, and accuracy. This fine-grained approach allows for a more detailed, objective, and comprehensive assessment of data quality, extending its applicability to complex industrial datasets.

Meanwhile, Vetro et al. [24] focused on open government data and identified key dimensions such as completeness, traceability, timeliness, and accuracy. They argue that open data must go beyond meeting technical standards to include considerations for usability and transparency. These elements are essential for enhancing public trust and engagement, and they provide a useful framework for assessing data quality in the context of public governance.

While there has been substantial research on data quality assessment frameworks across various domains such as transportation [25], healthcare [26], and construction [27], there remains a notable gap in the standardized scoring of government hotline data records. These data records represent a critical communication channel between the government and the public, and the quality of these data plays a pivotal role in the effectiveness of government services.

Government hotline data have specific characteristics and challenges that set them apart from other data types. For instance, they often contain large volumes of unstructured textual data, making their quality assessment particularly complex. Given these challenges, this study proposes a comprehensive data record standard evaluation system for government hotline data, focusing on dimensions such as reliability, correctness, and consistency. This system seeks to address the unique features of government hotline data and ensure that these records comply with standards, thereby enhancing government service efficiency and improving public satisfaction.

The process of evaluating government hotline data compliance involves several key steps, including topic identification, entity recognition, and text classification. Topic identification is crucial for extracting meaningful thematic information from large volumes of text. Traditional methods such as part-of-speech tagging (POS) and TF-IDF have been employed to identify topics within text [28]. However, models like Latent Dirichlet Allocation (LDA), though useful for long texts, often struggle with short texts due to limited context awareness [29,30]. To address these limitations, more recent approaches, such as the Adaptive Online Biterm Topic Model (AOBTM) [31], and the application of pre-trained models like BERT for topic classification [32], have shown promise in improving topic identification accuracy, especially in dynamic environments. The BERTopic model further refines BERT's capabilities for domain-specific topic recognition [33].

Entity recognition, another essential component of this evaluation system, involves identifying and categorizing named entities within text. Early methods were based on rule-based and dictionary-based approaches [34–38], but more recent advancements have leveraged machine learning [39,40] and deep learning models [41] such as LSTM and BERT to enhance the accuracy of entity recognition, especially in large and complex datasets.

Finally, text classification is a critical step in categorizing government hotline data into predefined categories. Common classification techniques include k-nearest neighbors, naive Bayes, support vector machines (SVMs), and, more recently, graph neural networks [42–44]. These methods contribute to the structured analysis of unstructured textual data, enabling more efficient processing and interpretation.

Quantitative methods for data quality assessment are essential for ensuring the reliability and validity of datasets across various domains. These methods use defined metrics to measure specific dimensions of data quality, enabling organizations to assess and enhance their data's fitness for use. Typically, quantitative assessments focus on several key dimensions, such as accuracy, completeness, consistency, timeliness, and uniqueness, each evaluated using specific metrics [45].

Accuracy involves comparing data to a known standard or ground truth. It ensures that data reflect the real-world phenomena they are supposed to represent. Completeness measures the proportion of available versus missing data entries, indicating whether

data are sufficiently detailed and comprehensive. Consistency ensures uniformity and the absence of contradictions within or across datasets, ensuring that data align with established rules or standards. Timeliness evaluates whether data are current and available when needed, which is crucial for time-sensitive applications such as decision-making in government services. Uniqueness identifies and eliminates duplicate records, which is especially important in maintaining clean and reliable datasets.

The diversity of proposed dimensions—more than 50 distinct dimensions in total—demonstrates the complexity and multifaceted nature of data quality assessment [45]. For example, Lewis et al. [46] systematically reviewed the factors affecting electronic health record data and proposed five core dimensions—completeness, correctness, consistency, reasonableness, and timeliness—as a foundational framework for structured and semi-structured data.

Building upon this, Song et al. [47] introduced a fine-grained data quality assessment framework, emphasizing sub-dimensions such as completeness, normativeness, consistency, uniqueness, and accuracy, offering more granular diagnostic capabilities for data quality analysis. However, these frameworks are often generalized, and they may not fully capture the unique challenges of textual and semi-structured data in government hotline work orders, which often contain free-text entries, inconsistent formats, and variable reporting standards.

Various frameworks have been developed to systematically assess data quality. Hierarchical data quality frameworks, for example, organize dimensions from general to specific, facilitating comprehensive evaluations. These are particularly suited for complex scenarios such as big data environments [48]. Systematic literature reviews (SLRs) have summarized practices and challenges in data quality assessment, highlighting the need for standardized methodologies and the integration of quality management into data processes [49]. Domain-specific frameworks, such as those used in public health, emphasize sector-specific data quality needs [50]. Despite the progress, challenges remain, particularly in managing the vast volume, velocity, and variety inherent in big data, the lack of universally accepted standards for data quality dimensions, and the critical need for high-quality data in machine learning applications [51].

However, while the literature has provided useful frameworks, government hotline data evaluation has mostly involved theoretical framework-based analyses, often lacking empirical studies to validate these models. These studies primarily focus on macro-level indicators like accuracy, completeness, consistency, and timeliness, but they seldom incorporate detailed or semantic indicators derived from text content. This is where our study innovates by focusing specifically on Chinese government hotline data, employing advanced methods such as BERT for precise text processing and text classification prediction, which significantly enhances the classification accuracy.

BERT (Bidirectional Encoder Representations from Transformers) has emerged as a leading method for natural language processing (NLP) tasks, particularly for handling unstructured text data. Its key advantage lies in its ability to capture the contextual relationships in text, allowing it to understand the nuances of language, which is crucial for processing government hotline work orders containing complex and varied textual data. Traditional NLP methods, such as rule-based or shallow machine learning models, often fail to grasp the deeper context or subtle patterns in the language, making BERT a more appropriate choice for this task.

The hybrid model BERT-BiLSTM-CRF integrates the strengths of BERT's contextualized embeddings with the BiLSTM (Bidirectional Long Short-Term Memory) network's ability to model sequential dependencies and CRF's (Conditional Random Field) capacity to enforce structured prediction constraints. This combination is particularly effective

for named entity recognition (NER), which is essential for extracting critical semantic information such as addresses, names, dates, and contact details from government hotline work orders.

The advantages of BERT-BiLSTM-CRF over traditional methods are as follows:

Contextual Understanding: BERT's pre-trained models enable deeper contextual understanding compared to traditional models like TF-IDF or Word2Vec, which treat words independently and fail to capture the context. For government hotline data, which often contain short, context-dependent phrases, BERT's ability to understand words in relation to their surrounding context ensures a higher level of precision and recall in extracting meaning from text.

Improved Classification Accuracy: Traditional methods like SVM (support vector machine) or naive Bayes focus on vectorizing text and using shallow models for classification. However, BERT integrates language understanding directly into the model, improving the classification performance significantly. In our study, incorporating BERT's language understanding improves the precision and recall for categorizing hotline data, leading to more accurate assessments of data compliance with predefined standards.

Enhanced Entity Recognition: Entity recognition is a critical task in government hotline work order texts, which often contain unstructured, free-text entries. Models like CRF or BiLSTM have been used for entity extraction but often struggle with semantic nuances in large datasets. By combining BiLSTM for sequential learning with CRF for structured prediction, our model is able to recognize locations, names, addresses, and other key entities with significantly higher accuracy compared to conventional methods, such as rule-based approaches or standalone machine learning techniques.

Adaptability and Fine-tuning: BERT offers a unique advantage in its adaptability. It can be fine-tuned for specific domains, such as government hotline data, using transfer learning. This approach allows for better performance on domain-specific tasks compared to traditional models, which would require extensive retraining for each new task or dataset. Our study demonstrates this advantage by fine-tuning BERT specifically on Chinese government hotline data, optimizing its performance for the task at hand.

While BERT has demonstrated clear advantages in this study, it is important to acknowledge the limitations of alternative methods. Traditional NLP methods like TF-IDF and naive Bayes focus on bag-of-words representations and often fail to capture the contextual meaning of words in short or ambiguous texts. SVM and Logistic Regression may perform well on structured datasets but struggle with unstructured text data like those found in government hotline work orders, which often contain inconsistent formats and varied vocabulary. Moreover, LSTM-based models, though capable of capturing sequential dependencies, do not have the sophisticated understanding of language that BERT provides, making them less effective in tasks involving complex textual content.

In contrast, BERT, by leveraging its transformer architecture, has revolutionized the processing of unstructured text by considering the entire context of a sentence rather than just individual words. This makes BERT particularly well suited for understanding the context-specific terminology and phrasing in government hotline work orders. The incorporation of BiLSTM-CRF further enhances the model by ensuring that the recognized entities maintain the correct sequential order, which is critical in the structured data extraction tasks required for evaluating government hotline data compliance.

2.2. Research Hypothesis

This study proposes a tailored data record standard evaluation system specifically designed for government hotline data. Drawing from the authoritative literature and the distinct characteristics of government work order systems, we define three core evaluation

dimensions—reliability, correctness, and consistency—each supported by sub-indicators with robust theoretical and empirical foundations.

Reliability assesses whether the data effectively support accurate decision-making and task routing. According to Batini et al. [52], data reliability is essential for ensuring that information can be effectively utilized for intended business operations. Similarly, Pipino et al. [45] emphasize that reliable data should consistently meet user expectations for decision-making tasks, which validates our inclusion of classification predictability (the clarity and accuracy of assigning problem categories) and dispatch accuracy (the precise routing of problems to appropriate departments) as key sub-indicators.

Correctness pertains to factual accuracy and compliance with data standards. Wang and Strong [53] highlight accuracy as a fundamental dimension reflecting the degree to which data correctly represent real-world conditions. Hamdy et al. [54] empirically confirm that record accuracy is directly correlated with overall data usability and quality, reinforcing our choice of record accuracy as a key sub-indicator. Additionally, Tian et al. [55] demonstrated that address accuracy directly impacts downstream logistics operations, further validating our inclusion of address accuracy as critical for government service delivery.

Consistency ensures internal coherence and logical integrity within textual records. Redman [56] established consistency as a critical dimension of data quality, referring to the uniformity and coherence of data within and across records. Building upon Gharehchopogh and Khalifelu's [57] work on evaluating topic relevance in textual data, we include adjacent sentence similarity and full-text similarity as indicators of logical consistency and minimal redundancy within hotline work orders, improving semantic integrity and operational efficiency.

Compared to previous studies primarily focused on structured data or general metrics, our framework uniquely addresses the challenges posed by unstructured or semi-structured textual records, thus providing a nuanced and operationally relevant evaluation system for government hotline data.

Numerous studies affirm the critical role of rigorous data quality standards in improving governmental performance. Li and Shang [58] showed that robust data standards significantly enhance administrative efficiency, while Cheng and Liu [59] demonstrated that improved data governance substantially enhances policy effectiveness. Consequently, we propose the following hypotheses:

H1: *Data record standards positively affect work order completion*.

Furthermore, the impact of data quality standards may vary across different work order fields. Studies indicate that sectors with structured workflows, like transportation and healthcare, typically exhibit greater sensitivity to enhanced data accuracy and standardization [60,61]. Hence, we hypothesize the following:

H2: *The work order field moderates the relationship between data record standards and work order completion*.

Similarly, organizational capabilities also influence how effectively data standards are utilized. Strong organizational competencies in data management significantly enhance data quality implementation and operational outcomes [62]. Departments with better training and clearer processes are more effective in utilizing standardized data, while those with limited capacities may struggle to realize these benefits. Thus, we propose the following:

H3: *The work order handling department moderates the relationship between data record standards and work order completion.*

These hypotheses form the basis of our proposed framework, as illustrated in Figure 2. This figure illustrates the evaluation framework for government hotline data record standards. In particular, six secondary indicators—including dispatch accuracy, classification predictability, record accuracy, address accuracy, adjacent sentence similarity, and full-text similarity—are employed to assess data standards. The evaluation results are further linked to key work order attributes, such as the work order field and handling department, in order to analyze their influence on work order completion.

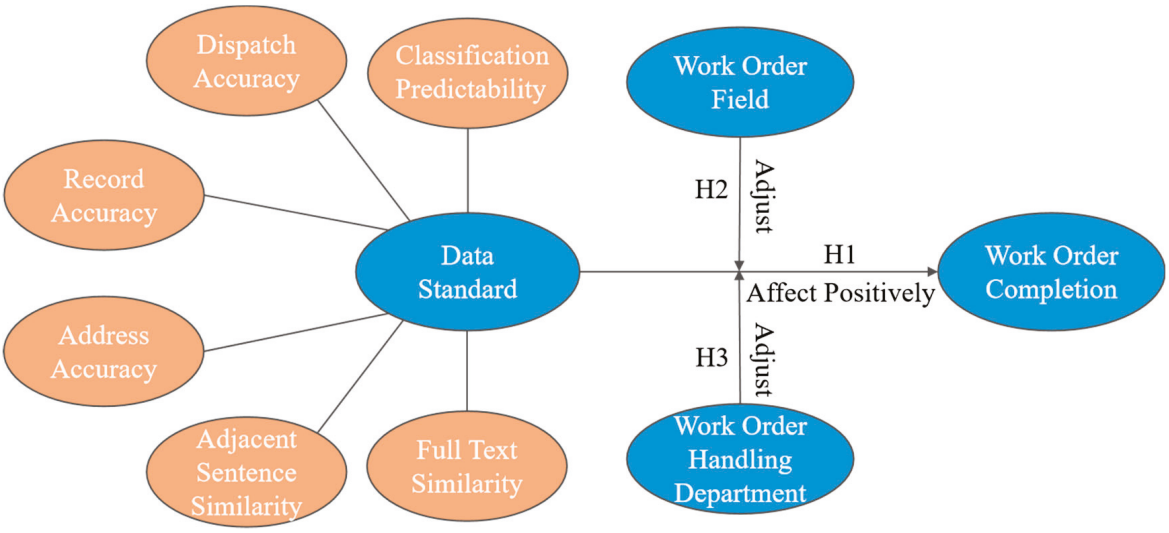


Figure 2. Framework for evaluation of data record standards and their impact on work order completion.

3. Research Design

3.1. Data Collection

The data for this study were sourced from a government hotline system, spanning from January 2019 to July 2021. After data cleaning and de-duplication, a total of 266,850 records were obtained.

3.2. Research Methods

To systematically analyze the government hotline data record standards, this study employed a series of computational methods and techniques. Specifically, six evaluation indicators were established to comprehensively assess the data quality in different dimensions, including classification predictability, dispatch accuracy, record correctness, address accuracy, adjacent sentence similarity, and full-text similarity. The data sources, calculation logic, and potential sources of error for each indicator were carefully considered in the analysis. Among these indicators, dispatch accuracy and address accuracy are defined as binary variables, serving as deterministic indicators with clear evaluation criteria. In contrast, classification predictability is measured using the output probability of a deep learning classification model, which helps to reduce uncertainty and provides a probabilistic recommendation of correctness. For the text similarity dimensions, including adjacent sentence similarity and full-text similarity, this study utilized a BERT-based semantic similarity model, which offers relatively high accuracy in capturing semantic relationships [63]. Furthermore, to mitigate potential deviations in the similarity calculations, both the average similarity and maximum similarity scores were adopted as evaluation metrics. The overall

analytical framework, along with the implementation steps of the proposed approach, is illustrated in Figure 3.

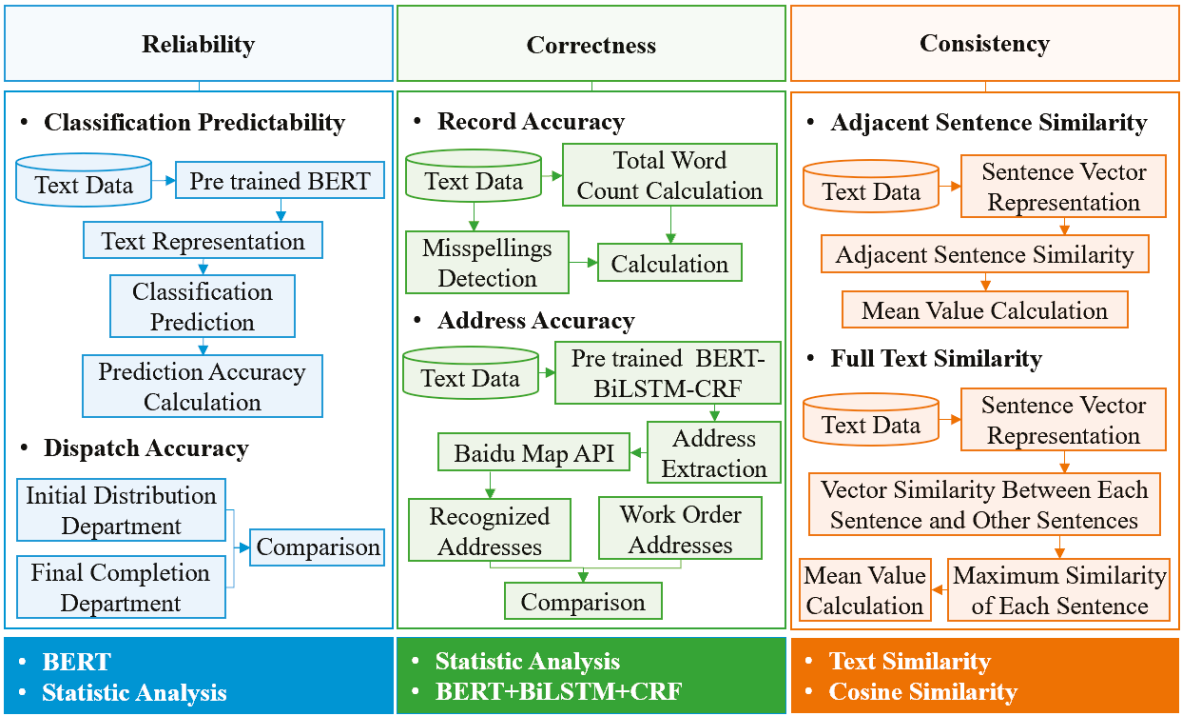


Figure 3. Technical route of the proposed framework.

3.2.1. Data Pre-Processing

For natural language processing classification algorithms—particularly BERT-based models—the input length is constrained to 512 characters, necessitating thorough pre-processing of the hotline texts. Given that the average length of government hotline texts is approximately 112 characters, we selected 128 characters as the maximum input length for the BERT model. The detailed data pre-processing steps were as follows:

1. Duplicate and Empty Entry Removal: We identified and removed duplicate records by comparing the textual content. Empty or missing entries were also systematically detected and excluded to ensure data integrity.
2. Missing Values and Outlier Handling: Entries containing missing textual fields or irrelevant special characters (such as random symbols or clearly erroneous input) were filtered out. Outliers—texts significantly deviating from normal hotline contents—were identified using statistical measures (e.g., text length distributions) and manually reviewed before exclusion.
3. Chinese Text Segmentation and Stopword Removal: The Jieba segmentation tool was employed for effective Chinese text segmentation, converting texts into token sequences. Subsequently, standard Chinese stopword lists were applied to remove frequently occurring but semantically irrelevant terms (e.g., auxiliary verbs, pronouns, common connectors).
4. Keyword Extraction and Text Length Adjustment: For texts exceeding the 128-character limit, we utilized the TF-IDF algorithm to rank terms based on their relevance and importance. High-ranking keywords were then selected to reconstruct concise texts that complied with the 128-character constraint, ensuring critical semantic content was preserved.
5. Denoising and Data Quality Enhancement: To enhance data quality, we implemented additional denoising techniques, including removing noisy patterns such as repeti-

tive text fragments, overly frequent boilerplate sentences, and irrelevant numerical strings. These steps significantly improved the accuracy and reliability of subsequent classification and analysis tasks.

No specialized pre-processing was required for the named entity recognition (NER) or semantic similarity calculation algorithms. Specifically, NER involves annotation of the entire original text, while semantic similarity calculations segment texts into individual sentences, each typically shorter than 128 characters.

3.2.2. Calculation Method for Classification Predictability

The secondary indicator for reliability that determines the category of work orders is classification predictability. This indicator is primarily determined based on the BERT model for classification. BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained deep learning model that captures contextual information through a bidirectional transformer architecture, improving accuracy in natural language processing tasks. The principle of text classification using the BERT model is shown in Figure 4.

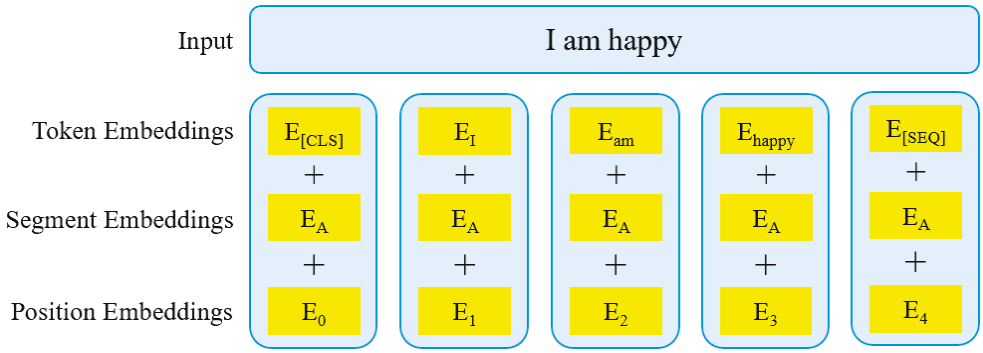


Figure 4. BERT model text processing principle.

It should be noted that due to the significant population and service disparities across different regions and departments, the volume of government hotline data is highly imbalanced. For instance, by the end of 2022, District M of City B had a resident population of approximately 980,000, whereas District N had only 320,000. Similarly, the scope of services provided by different government departments varies significantly. The urban management department handles issues such as street cleaning, noise pollution, and illegal construction, while the education department deals with school admissions and education resource allocation. Additionally, citizens' demands may shift based on specific events, such as the surge in hotline calls to the health department during the COVID-19 pandemic. This results in considerable data volume disparities across regions and departments, with some having very few work orders. Figure 5 illustrates the number of work orders closed by various districts and departments in City A in 2023.

Given this imbalance, a method was adopted to address the issue of misclassification due to insufficient data. For categories with fewer than 100 work orders, data augmentation was used to increase the number to 100. This augmentation was achieved using the Qwen-7B-Chat model, which was selected for the following reasons:

In August 2023, Alibaba Cloud announced the open-sourcing of two models from its Qwen series—Qwen-7B and Qwen-7B-Chat—available on both ModelScope and Hugging Face. Both models have a parameter scale of 7 billion. Qwen-7B, a decoder-only language model based on a transformer architecture, is similar to the LLaMA series. It was pre-trained on over 2.2 trillion tokens and supports a context length of up to 2048. In various benchmarks tested by Alibaba Cloud, Qwen-7B outperformed most existing models of

similar size. Table 1 compares the accuracy of Qwen-7B and other pre-trained language models on the C-Eval validation set.

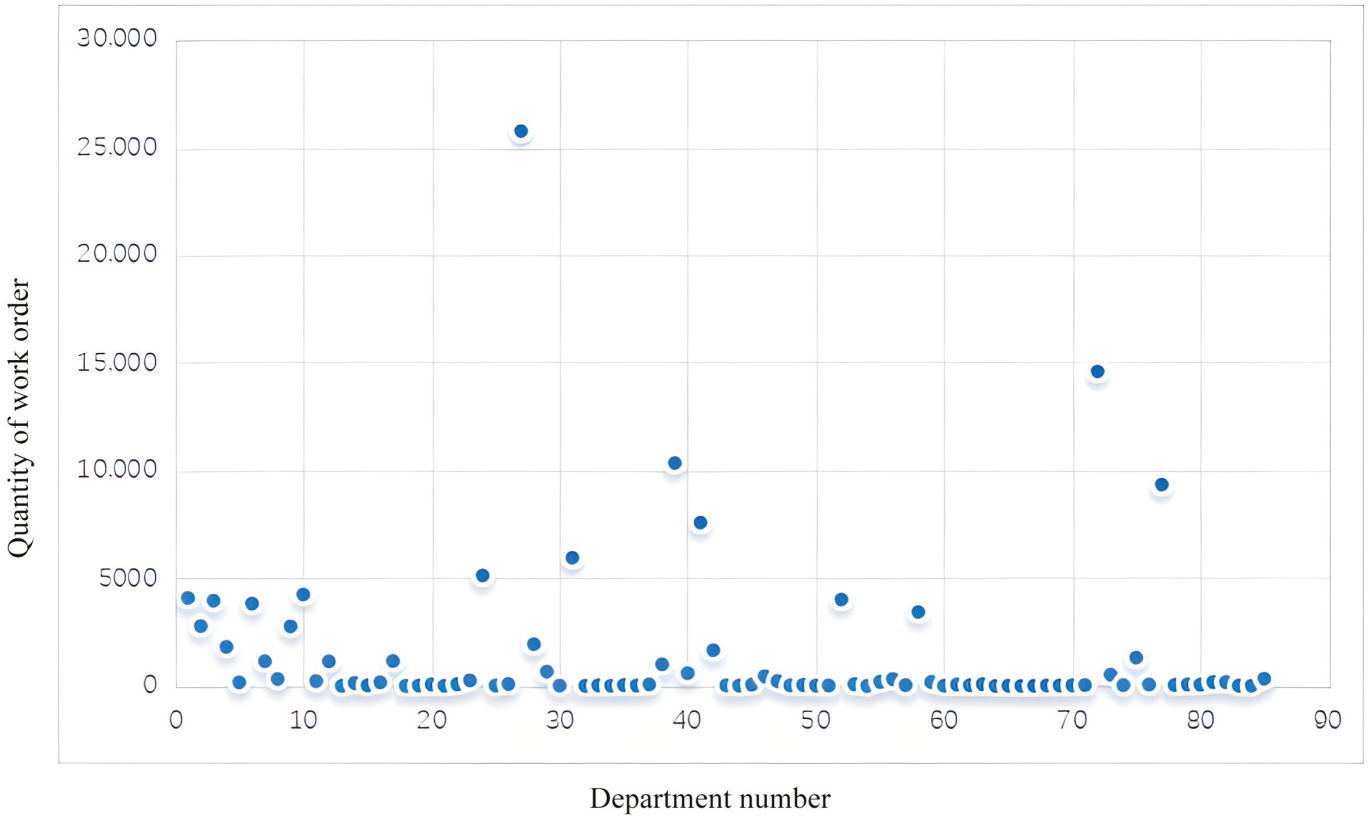


Figure 5. Number of work orders closed by various districts and departments in City A in 2023.

Table 1. Accuracy of Qwen-7B and other pre-trained language models.

Model	Average
Alpaca-7B	28.9
Vicuna-7B	31.2
ChatGLM-6B	37.1
Baichuan-7B	42.7
ChatGLM2-6B	50.9
InternLM-7B	53.4
ChatGPT-3.5	53.5
Claude-v1.3	55.5
Qwen-7B	60.8

The Qwen-7B model outperformed other models of similar size and even surpassed larger models in several cases. Its performance on the C-Eval test set compared to other models is shown in Table 2.

As can be seen from the table, Qwen-7B outperformed similar models and even exceeded some larger models. Therefore, in this study, we utilized Qwen-7B to enhance the data by generating additional work orders for categories with fewer than 100 cases. This was achieved by constructing prompt templates and invoking Qwen-7B's API, which reformulated the work order text using different vocabulary and structures while maintaining the original meaning. The prompt used for this process was "*Use the Qwen model to rewrite this sentence with different vocabulary and structure while strictly preserving the original meaning*".

Table 2. Performance of Qwen-7B and other models on C-Eval test set.

Model	Avg.	Avg. (Hard)	STEM	Social Sciences	Humanities	Others
ChatGLM-6B	38.9	29.2	33.3	48.3	41.3	38.0
Chinese-Alpaca-Plus-13B	41.5	30.5	36.6	49.7	43.1	41.2
Baichuan-7B	42.8	31.5	38.2	52.0	46.2	39.3
WestlakeLM-19B	44.6	34.9	41.6	51.0	44.3	44.5
AndesLM-13B	46.0	29.7	38.1	61.0	51.0	41.9
BatGPT-15B-sirius	47.0	31.9	42.7	57.5	48.6	43.6
ChatGLM2-6B	51.7	37.1	48.6	60.5	51.3	49.8
InternLM-7B	52.8	37.1	48.0	67.4	55.4	45.8
Baichuan-13B	53.6	36.7	47.0	66.8	57.3	49.8
Claude-v1.3	54.2	39.0	51.9	61.7	52.1	53.7
ChatGPT-3.5	54.4	41.4	52.9	61.8	50.9	53.6
Qwen-7B	59.6	41.0	52.8	74.1	63.1	55.2

The specific calculation process for classification predictability involves tokenizing and vectorizing the text data using a fine-tuned pre-trained BERT model and then inputting the processed text into the BERT model to obtain classification results. The formula is as follows:

$$R_c = P(y_i|x_i) = softmax(BERT(x_i))$$

where R_c represents the classification predictability and $P(y_i|x_i)$ denotes the predicted classification probability of the input text x_i.

3.2.3. Calculation Method for Dispatch Accuracy

The secondary indicator for reliability in evaluating whether the dispatched department for a government work order is accurate is dispatch accuracy, which is calculated by comparing the department recommended by the classification model to the final handling department. The calculation process is illustrated in Figure 6.

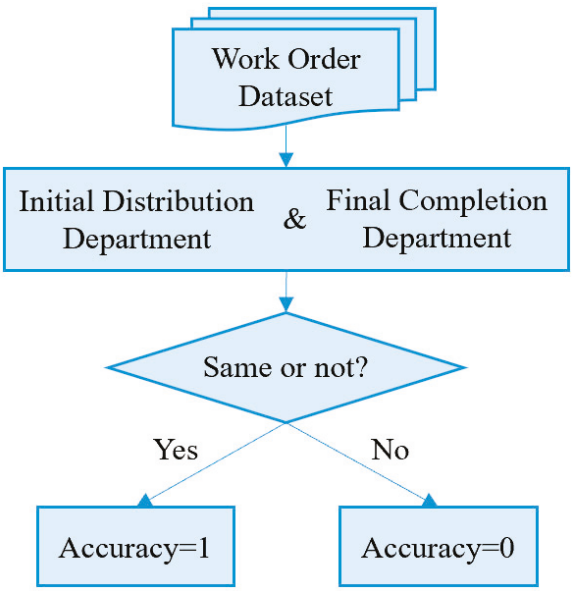


Figure 6. Flowchart of dispatch accuracy calculation.

The formula for calculating this indicator is as follows:

$$R_d = \frac{\sum_{i=1}^{n} 1(y_i = \hat{y}_i)}{n}$$

where R_d represents the accuracy of dispatching, y_i is the actual handling department, $\hat{y}_i$ is the department predicted by the model, 1 $(y_i = \hat{y}_i)$ is an indicator function that is equal to 1 if $y_i = \hat{y}_i$ and 0 otherwise, and n is the total number of work orders.

3.2.4. Calculation Method for Record Accuracy

Record accuracy is a secondary indicator under the primary indicator of correctness. This indicator measures whether the textual records in each work order are accurate and whether there are spelling or grammatical errors. Spelling error detection in work order texts is performed using the 'pycorrector' package in Python3.11, which uses language model perplexity (PPL) to determine the correctness of a character. The spelling error detection process is shown in Figure 7.

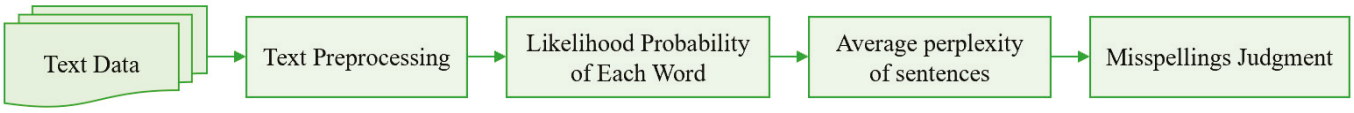


Figure 7. Spelling error detection process.

The record accuracy indicator is finally assessed by calculating the proportion of non-spelling errors relative to the total number of characters. The formula is as follows:

$$C_w = 1 - \frac{E}{W}$$

where C_w represents the accuracy of records, E denotes the number of spelling errors, and W is the total number of characters.

3.2.5. Calculation Method for Address Accuracy

Another secondary indicator under the correctness dimension is address accuracy, which is designed to evaluate the correctness and consistency of address information recorded in government hotline work orders. Specifically, this indicator measures whether the address information extracted from the textual description of a work order matches the structured address field provided in the dataset. To achieve this, a named entity recognition (NER) method was applied to automatically extract address entities from the unstructured text content of each work order. The NER process is based on a BERT-BiLSTM-CRF model, which effectively integrates three advanced techniques: (1) the BERT model is used to capture rich contextual semantic features of the input text; (2) the BiLSTM (Bidirectional Long Short-Term Memory) network is employed to learn sequential dependencies in both the forward and backward directions; and (3) the CRF (Conditional Random Field) layer is utilized to optimize the sequence labeling results and ensure the accuracy and integrity of address entity recognition. The workflow and principle of this address extraction model are illustrated in Figure 8. After extraction of the address entity, the identified address is further standardized and verified by querying the Baidu Maps API to determine its administrative region or geographic location. The obtained address information is then compared with the structured address field in the original work order dataset. If the two address sources correspond to the same administrative region or location, the address accuracy is recorded as 1 (correct); otherwise, it is recorded as 0 (incorrect). This binary evaluation criterion ensures the precision of the address accuracy indicator and enables quantitative assessment of the correctness of address information in the hotline data records.

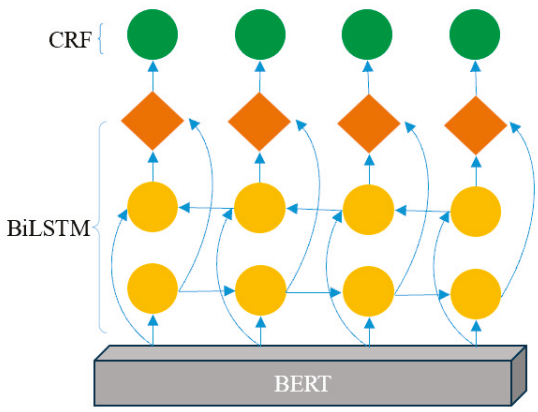


Figure 8. Working principle of the BERT-BiLSTM-CRF model.

3.2.6. Calculation Method for Theme Similarity

The primary indicator for consistency in the evaluation of hotline data record normativity mainly refers to thematic consistency, including the similarity between adjacent sentences (adjacent sentence similarity) and the similarity of each sentence to the entire text (full-text similarity). The calculation of these indicators is based on a BERT-based semantic similarity model, which leverages the deep contextual understanding capabilities of BERT to measure the semantic closeness between sentences.

The calculation process is carried out as follows: First, each pair of sentences to be compared is input into a pre-trained BERT model, which encodes the sentences into dense vector representations within a high-dimensional semantic space. The BERT model captures both the syntactic structure and contextual semantics of the sentences, ensuring that the generated vectors accurately reflect their meanings. Subsequently, the cosine similarity between the two sentence vectors is computed to quantify their semantic similarity. The cosine similarity score ranges from 0 to 1, where a higher score indicates a higher degree of semantic consistency between sentences.

1. Adjacent Sentence Similarity: This secondary indicator uses BERT to convert sentences into vectors, calculates the similarity between adjacent sentence vectors, and then takes the average of all similarities. The formula is as follows:

$$S_a = \frac{1}{n-1}\sum_{i=1}^{n-1}\cos(V_{s_i}, V_{s_{i+1}})$$

where V_{s_i} represents the vector of the *ith* sentence, cos is the cosine similarity measure, and n is the total number of sentences. The calculation method for adjacent sentence similarity is further illustrated in Figure 9.

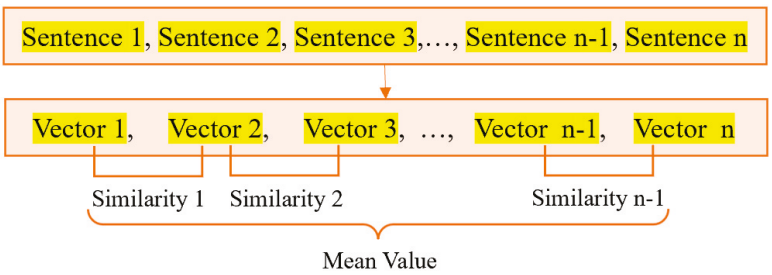


Figure 9. Calculation method for adjacent sentence similarity.

2. Full-Text Similarity: This secondary indicator first converts sentences into vectors using BERT, then calculates the maximum similarity between each sentence and all

other sentences, and finally averages these maximum similarities. The formula is as follows:

$$S_o = \frac{1}{n}\sum_{i=1}^{n} \max_{j \neq i} \cos(V_{s_i}, V_{s_j})$$

where V_{s_i} represents the vector of the ith sentence, cos is the cosine similarity measure, and n is the total number of sentences. The full-text similarity calculation differs from that for adjacent sentence similarity in that it calculates the maximum similarity of each sentence with respect to the other $n - 1$ sentences and then averages these maximum similarities. The calculation process is illustrated in Figure 10.

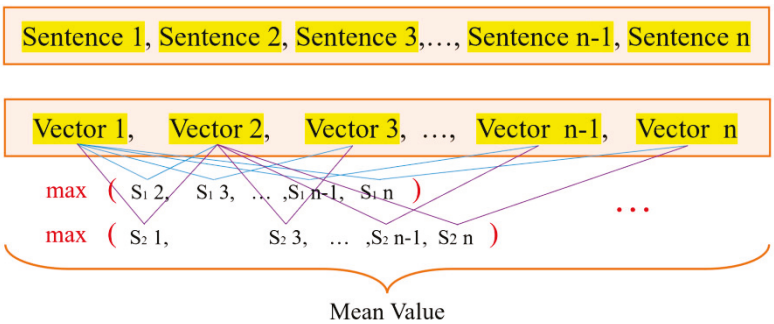


Figure 10. Calculation method for full-text similarity.

4. Experimental Results

4.1. Descriptive Statistics

To facilitate quantitative analysis, we encoded the work order fields and handling departments. The work order fields were coded from 0 to 19, covering various areas such as transportation and housing (Table 3), while the handling departments were categorized into district-level departments (0), municipal-level departments (1), and state-owned enterprises (2).

Table 3. Coding for different work order fields.

Category	Code
Transportation	0
Housing Security	1
Party and Government Affairs	2
Public Safety	3
Public Services	4
Agriculture, Forestry, Water, and Land	5
Healthcare	6
Urban and Rural Construction	7
Urban Comprehensive Management	8
Safety Supervision	9
Government Administration	10
Livelihood Security	11
Public Security and Legal Affairs	12
Environmental Protection	13
Science, Education, Culture, and Health	14
Science, Education, Culture, and Tourism	15
Comprehensive Economic Management	16
Mass Organizations	17
Industry Supervision	18
Outside Scope of Acceptance	19

This study analyzed hotline data record standards using six research indicators: classification predictability, dispatch accuracy, record accuracy, address accuracy, adjacent sentence similarity, and full-text similarity. Table 4 presents the minimum, maximum, variance, mean, and number of observations for each indicator.

Table 4. Descriptive statistics for the utilized indicators.

Indicator	Minimum	Maximum	Variance	Mean	Observations
Classification Predictability (CP)	0.037	0.999	0.043	0.612	266,850
Dispatch Accuracy (DA)	0	1	0.048	0.950	266,850
Record Accuracy (RA)	0.893	1	4.08×10^{-5}	0.997	266,850
Address Accuracy (AA)	0	1	0.237	0.375	266,850
Adjacent Sentence Similarity (ASS)	0	0.936	0.006	0.745	266,850
Full-Text Similarity (FTS)	0	1	0.008	0.814	266,850
Work Order Field (WOF)	0	19	20.799	7.923	266,850
Work Order Handling Department (WOHD)	0	2	0.337	0.366	266,850
The Completion of the Work Orders (CWO)	0	1	0.137	0.836	266,850

From Table 3, it can be seen that the average values and variances of the indicators reflect the performance of the data in various aspects. Classification predictability and dispatch accuracy had relatively high average values, indicating that most work orders had good classification and dispatch accuracy. The average value of record accuracy was close to 1, suggesting that there were few spelling errors in the work order records. The variance value of address accuracy was larger, indicating some fluctuation in address records. The average values for adjacent sentence similarity and full-text similarity were high, indicating a high level of consistency in adjacent sentences and overall semantics in the work order texts. For cases where the similarity is 0, we need to provide a separate explanation; this occurs when the sentence is too short to calculate its similarity with other sentences, and such brief descriptions often result in reduced completeness due to insufficient information.

4.2. Experimental Results of Deep Learning Models

Table 5 outlines the hyperparameters and dataset sizes for both text classification and named entity recognition (NER) tasks. Specifically, for text classification, the dataset sizes (train, test, and dev) represent the number of individual text samples. In contrast, for the named entity recognition task, the dataset sizes indicate the total number of characters processed. Additionally, the maximum sequence length (max_seq_length) for text classification was set to 128, while that for named entity recognition was set to 256, reflecting the differing requirements of these two deep learning tasks. Other parameters (e.g., batch size, learning rate, and epochs) are also task-specific and were optimized accordingly.

Table 5. Experimental parameters of the algorithms.

Parameters	Text Classification	Named Entity Recognition
max_seq_length	128	256
train_epoches	1	20
batch_size	128	4
learning_rate	5×10^{-5}	3×10^{-5}
rnn_dim		128
train_data_size	871,109	1,734,751
test_data_size	48,396	257,841
dev_data_size	48,397	243,060

Table 6 presents the evaluation results of the proposed models in two natural language processing tasks: text classification and named entity recognition (NER). For the text

classification task, the model achieved consistently high performance across all metrics, with an accuracy of 0.9219, precision of 0.9446, recall of 0.9219, and F1-score of 0.9234, demonstrating its effectiveness in distinguishing text categories. For the NER task, the model obtained an accuracy of 0.9548, with a precision of 0.7048, a recall of 0.7500, and an F1-score of 0.7267. These results indicate that the model is capable of accurately identifying named entities while maintaining a balanced performance between precision and recall.

Table 6. Experimental results obtained for the considered algorithms.

Algorithms	Accuracy	Precision	Recall	F1
Text classification	0.9219	0.9446	0.9219	0.9234
Named entity recognition	0.9548	0.7048	0.7500	0.7267

4.3. Data Standardization

To ensure that the indicators are comparable and compute the data record normativity score, we standardized the data using the 'StandardScaler' method from the sklearn package in Python. 'StandardScaler' removes the mean of each feature and scales it to unit variance, allowing data from different features to be compared on the same scale. The standardized data follow a standard normal distribution, with a mean of 0 and a standard deviation of 1.

Using this method, we standardized the six research indicators, enabling their comparison on the same scale. Standardized data provide a more accurate reflection of each indicator's contribution to the hotline data record normativity score.

4.4. Calculation of Total Score

To comprehensively evaluate the data record standard of each work order, we calculated the total score through summation of the six standardized indicators. The specific formula is as follows:

$$S = S_{R_c} + S_{R_d} + S_{C_w} + S_{C_a} + S_{S_a} + S_{S_o}$$

where S is the score of data record standard, S_{R_c} is the classification predictability after standardization, S_{R_d} is the accuracy of dispatch after standardization, S_{C_w} is the correctness of records after standardization, S_{C_a} is the correctness of addresses after standardization, S_{S_a} is the similarity of adjacent sentences after standardization, and S_{S_o} is the similarity of the entire text after standardization.

4.5. Linear Regression and Moderating Effects

In this study, we utilized a multiple linear regression model to explore the relationships between data record scores and work order completion rates, and we further examined the moderating effects of the work order field and the department handling the work order. The completion of work orders was considered as the dependent variable (Y), while the data record score, work order field, and handling department were considered as independent variables (X); furthermore, b represents the effect of each independent variable on the outcome, while u captures the error or variation in the outcome not explained by the model. In the moderating effect analysis, the work order field and handling department were treated as moderating variables interacting with the data record score in order to assess their influence on the completion rate.

$$y = b_0 + b_1x_1 + b_2x_2 +, ..., +b_nx_n + u$$

The moderating effects were assessed by analyzing the significance and direction of the interaction terms relative to the main effects. These effects can be categorized into four scenarios based on the significance and sign of the interaction term, as summarized in Table 7.

Table 7. Four types of moderating effects.

Main Effect Sign	Interaction Term Sign
Significant positive effect of the independent variable on the dependent variable	Interaction term (independent variable× moderating variable) is significantly positive
	Interaction term (independent variable× moderating variable) is significantly negative
Significant negative effect of the independent variable on the dependent variable	Interaction term (independent variable× moderating variable) is significantly positive
	Interaction term (independent variable× moderating variable) is significantly negative

4.6. Empirical Analysis

A correlation analysis was conducted using the cor() function in R, the results of which are presented in Figure 11, showing the correlation matrix of the key variables. As illustrated in the matrix, there was a strong linear correlation (coefficient = 0.90) between "adjacent sentence similarity" (ASS) and "full-text similarity" (FTS). Considering the risk of multicollinearity due to this high correlation, the variable "adjacent sentence similarity" was removed from the subsequent multiple linear regression analysis, thus ensuring the robustness and reliability of the model results.

Figure 11. Linear correlation matrix of variables.

From Table 8, it can be observed that dispatch accuracy and record accuracy were negatively correlated with work order completion, but the correlation between record accuracy and work order completion was not statistically significant (0.08). The statistical description of the data shows that the mean value of record accuracy was very close to 1, and the variance was very small, indicating that the accuracy of the textual descriptions in the work order data was very high. As a result, its effect on work order completion was not significant. Similarly, the mean value of dispatch accuracy was also very close to 1, suggesting that there were too many positive samples in the data, potentially affecting its correlation with work order completion. Intuitively, it would be expected that as dispatch accuracy increases, work order completion should also increase.

Table 8. The regression analysis results of five indicators with respect to work order completion.

	Estimate	**Std. Error**	***t* Value**	**Pr (>|t|)**
Classification Predictability	0.080977	0.003441	23.534	$<2 \times 10^{-16}$ ***
Dispatch Accuracy	−0.029718	0.003253	−9.134	$<2 \times 10^{-16}$ ***
Record Accuracy	−0.197684	0.112459	−1.758	0.08
Address Accuracy	0.017291	0.001475	11.725	$<2 \times 10^{-16}$ ***
Full-Text Similarity	0.811878	0.018114	44.820	$<2 \times 10^{-16}$ ***

Significance *p*-value: ***, <0.001.

The remaining three indicators—address accuracy, full-text similarity, and classification predictability—showed significant positive correlations with work order completion, aligning well with our expectations. However, dispatch accuracy and record accuracy exhibited exceptionally high average values (0.95 and 0.997, respectively), indicating substantial data imbalance. Such an imbalance can introduce counterintuitive correlations, which underscores the necessity and appropriateness of using the integrated data quality scoring model discussed previously. Additionally, considering the high collinearity between adjacent sentence similarity and full-text similarity, we included both indicators within the integrated score, which helps to mitigate uncertainties arising from imbalanced data.

In the regression analysis, we examined the direct effect of integrated data record scores on work order completion rates. As shown in Table 9, there was a significant positive association between data record scores and work order completion ($p < 0.001$). This result aligns with previous findings emphasizing the importance of data standards in enhancing government governance; for instance, Huang [64] identified a significant moderating effect of data standards on government performance, while Niu et al. [65] similarly demonstrated a positive influence of data standards on environmental governance. Additionally, our regression analysis indicated that the work order field and handling department significantly affected work order completion rates.

Table 9. The results after adjusting the standard errors using the heteroskedasticity-consistent covariance matrix estimator.

	Estimate	**Std. Error**	***t* Value**	**Pr (>|t|)**
Classification Predictability	0.080977	0.003185	25.42	$<2 \times 10^{-16}$ ***
Dispatch Accuracy	−0.029718	0.002979	−9.975	$<2 \times 10^{-16}$ ***
Record Accuracy	−0.197684	0.114634	−1.724	0.085
Address Accuracy	0.017291	0.001458	11.863	$<2 \times 10^{-16}$ ***
Full-Text Similarity	0.811878	0.019171	42.350	$<2 \times 10^{-16}$ ***

Significance *p*-value: ***, <0.001.

To assess the robustness of the regression results, this study applied heteroskedasticity-robust standard errors. Specifically, Table 8 reports the baseline regression results obtained using the ordinary least squares (OLS) method, while Table 10 presents the results after adjusting the standard errors using the heteroskedasticity-consistent covariance matrix estimator (HC1). As can be seen from the two tables, the estimated coefficients and their significance levels remained largely consistent before and after the adjustment. This indicates that the presence of heteroskedasticity does not materially affect the estimation results, and the model exhibits a high degree of robustness.

Table 10. The regression analysis results of three indicators with work order completion.

	Estimate	**Std. Error**	***t* Value**	**Pr (>|t|)**
Score	0.0065431	0.0002627	24.90	$<2 \times 10^{-16}$ ***
Work Order Field	0.0024480	0.0001566	15.63	$<2 \times 10^{-16}$ ***
Work Order Handling Department	0.0583717	0.0012363	47.21	$<2 \times 10^{-16}$ ***

Significance *p*-value: ***, <0.001.

Next, we analyzed the moderating effect of the work order field (Table 11). The results showed that when the work order field was treated as a moderating variable, it significantly enhanced the positive effect of the data record score on the completion rate (Pr < 0.001). This suggests that, in certain fields, higher data quality leads to better work order completion rates, especially for specific types of work orders where the impact is more pronounced.

Table 11. Analysis of moderating effects by work order field.

	Estimate	**Std. Error**	***t* Value**	**Pr (>|t|)**
Score	0.0025635	0.0005676	4.516	6.29×10^{-6} ***
Work Order Field	0.0026238	0.0001573	16.676	$<2 \times 10^{-16}$ ***
Score × Work Order Field	0.0003299	0.0000615	5.364	8.16×10^{-8} ***

Significance *p*-value: ***, <0.001.

Conversely, the results regarding the moderating effect of the handling department (Table 12) showed that it weakened the positive impact of the data record score on the completion rate (Pr < 0.001). This implies that, in certain departments, the positive influence of data record scores on completion rates is reduced, possibly due to issues in their processes or data standardization practices that prevent the quality of data from fully enhancing the completion rate.

Table 12. Analysis of moderating effects by work order handling department.

	Estimate	**Std. Error**	***t* Value**	**Pr (>|t|)**
Score	0.0075021	0.0003232	23.213	$<2 \times 10^{-16}$ ***
Work Order Handling Department	0.0587536	0.0012369	47.499	$<2 \times 10^{-16}$ ***
Score × Work Order Handling Department	−0.0020442	0.0004363	−4.685	2.8×10^{-6} ***

Significance *p*-value: ***, <0.001.

These analyses indicate that data record scores have a significant impact on work order completion rates, with the work order field enhancing this effect and the handling department somewhat diminishing it. Based on these findings, we further analyzed the average data record scores across different work order fields. Figure 12 shows the average scores for different work order fields.

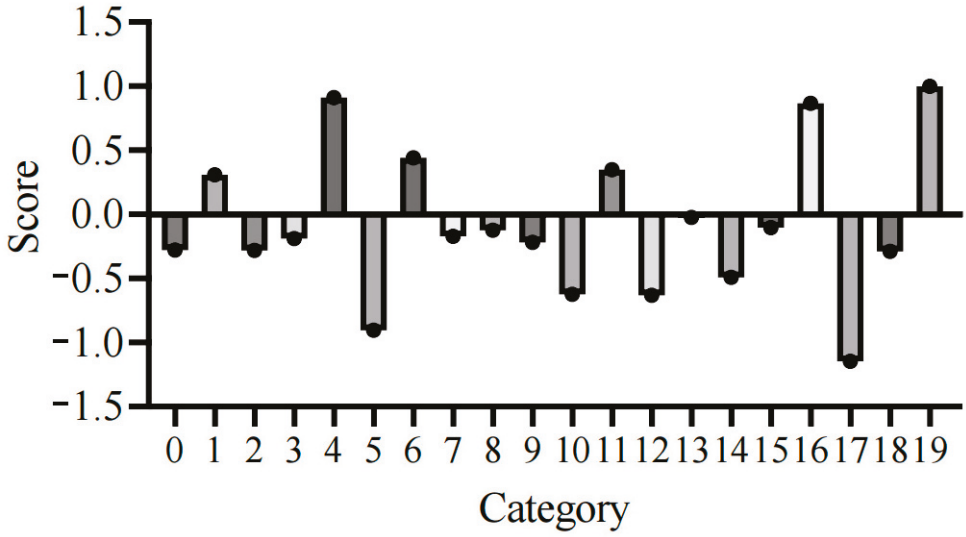


Figure 12. Statistical distribution of work order field scores across different categories.

As shown in the figure, there were significant differences in the average data record scores across different fields. Certain fields (e.g., fields 4, 16, and 19) had higher scores, indicating that work orders in these fields exhibit a higher degree of standardization and compliance with data record norms. In contrast, fields 5 and 17 had lower scores, suggesting that work order data quality in these areas needs improvement. These findings provide empirical evidence for the optimization of data record standards and allow for recommendations to be provided to government departments in order to improve their operations.

To further evaluate the importance of specific indicators affecting work order completion, a random forest algorithm was applied. The importance of various indicators was assessed, and the results are shown in Figure 13.

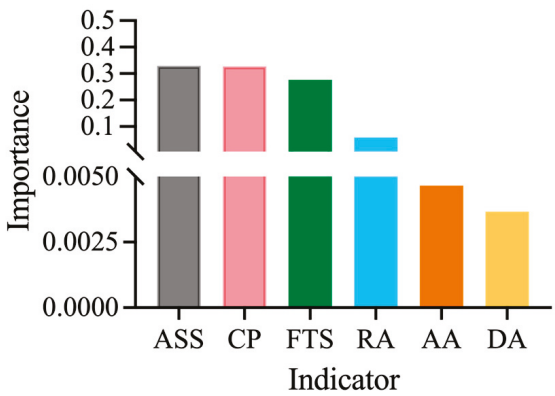


Figure 13. Evaluation of indicator importance.

The results presented in Figure 12 indicate considerable variability in the importance of individual indicators, as assessed by the random forest algorithm. It is crucial to note that the importance derived from this algorithm serves as a reference rather than an absolute measure, as random forest node splits tend to attribute higher predictive importance to indicators with greater data diversity and continuity. Specifically, semantic-related indicators such as adjacent sentence similarity (0.329), classification predictability (0.327), and full-text similarity (0.277), all of which are continuous data types, naturally exhibited higher importance due to their broader variability. Conversely, accuracy-related indicators, including address accuracy (0.005) and dispatch accuracy (0.004), are binary discrete data types and consequently yielded lower predictive significance within the random forest

framework. Although our integrated scoring model employed equal weights to mitigate potential biases due to imbalanced data distributions, these findings highlight intrinsic differences in the predictive capabilities of the indicators. Therefore, future strategies for data quality enhancement should carefully consider these characteristics, potentially adopting weighted integration methods informed by the relative predictive importance and nature (e.g., data type) of each indicator.

4.7. Case Analysis

To further illustrate the significance of data record standardization, we analyzed specific cases. Examples of work orders, their corresponding scores, and an analysis of the key issues that contributed to their respective scores are presented in Table 13.

Table 13. Specific case examples.

Case No.	Work Order Description	Score	Issue Analysis
1	The caller wants to drive a bus with a llama to XX for a performance but is unsure if animal quarantine procedures are required.	−7.47	The caller's area was not clarified, and quarantine procedures need to be handled in the corresponding jurisdiction. This caused a delay in work order completion.
2	The caller reports illegal parking on an unmarked cement road at XX Road, located at the intersection of XX, XX, and XX areas.	−14.96	The key term "illegal parking" was mistakenly written as "false parking". As machine learning algorithms are used for text classification, the misspelling of this keyword greatly impacted classification, reducing overall work order completion.

In Case 1, the data record standard score was −7.47, which was relatively low due to the caller's failure to provide clear regional information, resulting in dispatch errors and delayed completion of the work order. This case represents a typical scenario where insufficient address details lead directly to incorrect dispatching. Given that approximately 50% of work orders are processed by regional jurisdictions, address accuracy emerges as a critical indicator that is significantly correlated with work order completion rates. Although address accuracy—being discrete data—exhibited relatively low predictive importance within the random forest model, its substantial impact observed in real-world analyses validates its practical importance. This underscores the appropriateness and validity of the indicators selected in this study.

In Case 2, the standardization score was −14.96, the lowest among the examples presented. The critical issue was the misspelling of the key term "illegal parking" as "false parking", severely impacting the processing efficiency of the work order. This case underscores the significance of accurate recording in data standardization, as spelling errors prevent machine learning algorithms from accurately classifying issues, directly impairing the efficiency and accuracy of work order handling. Furthermore, as the semantic similarity calculations employed BERT-based sentence embeddings, even minor errors in individual words significantly affected the semantic representation of the entire sentence. Consequently, the observed spelling error notably reduced the semantic similarity score, contributing to an overall decline in the standardization score and ultimately leading to the work order's incomplete status.

Based on the above analyses, this study provides practical insights into hotline work order management within China's specific context. Given the complexity and variability inherent in government hotline data, indicators such as address accuracy and record accuracy play critical roles despite their lower predictive significance in random forest models. Address accuracy remains pivotal due to China's administrative structure, where around 50% of work orders are typically handled at the local jurisdiction level, emphasizing

the necessity of accurate regional information. Additionally, the prevalence of manual data entry and linguistic nuances, exemplified by issues such as misspellings, directly impact semantic clarity and subsequent automated classification using advanced NLP techniques like BERT embeddings. Therefore, to enhance the effectiveness of hotline operations, Chinese government departments should prioritize improving address details and textual accuracy, supplementing machine learning insights with real-world case analyses to comprehensively optimize service delivery and operational efficiency.

5. Conclusions and Recommendations

5.1. Conclusions

This study analyzed the relationship between government hotline data record standard scores and work order completion rates, revealing a significant positive correlation. Higher overall data quality scores, calculated by summing the six key evaluation indicators, were consistently associated with higher work order completion rates. Among these indicators, adjacent sentence similarity, classification predictability, and full-text similarity demonstrated the most substantial impact, underscoring the critical role of semantic clarity and structured coherence in data records. In contrast, record accuracy, address accuracy, and dispatch accuracy exhibited relatively lower predictive importance.

To further interpret these findings, this study incorporates relevant theoretical perspectives from data quality and information management. From the perspective of data quality, factors such as semantic coherence, contextual relevance, and logical consistency are essential for ensuring that information can be effectively understood and utilized in decision-making processes. The strong influence of adjacent sentence similarity, classification predictability, and full-text similarity reflects the fact that clear and well-structured data records facilitate accurate comprehension and task execution by frontline personnel, thereby contributing to higher completion rates.

Conversely, the relatively limited impact of record accuracy, address accuracy, and dispatch accuracy can be attributed to the diverse and complex nature of certain work order categories. In practice, work orders related to areas such as social governance, public facilities, and social security often involve ambiguous, unstructured, or incomplete information. In these cases, strict accuracy in individual data elements may not be the determining factor for successful resolution. Instead, overall semantic clarity and contextual integrity are more influential.

In summary, the differentiated impact of the six evaluation indicators reflects the inherent tension between data standardization and service flexibility in public governance. These findings not only validate the effectiveness of the proposed evaluation system but also provide a theoretical explanation for the varying influence of different data quality dimensions, offering valuable insights for optimizing hotline data management and improving service efficiency.

5.2. Recommendations

To further improve the quality and standardization of hotline work order data and enhance work order completion rates, this study offers the following recommendations for government departments:

(1) Strengthen Employee Training and Guidance: Although this study did not directly evaluate human input behaviors, the case analysis revealed that certain incomplete or inaccurately worded records—such as missing address details or incorrect keywords—may reduce the effectiveness of automated classification and dispatch. These issues suggest the potential value of targeted employee training as a complementary measure to improve the overall quality of data. Government departments

should implement tiered and role-specific training programs, including regular sessions for frontline staff in identifying key information, using standardized terms, and avoiding common errors. Real-world case studies should be incorporated to enhance practical skills. For reviewers and supervisors, training should focus on error correction, quality monitoring, and the delivery of feedback. The effectiveness of training should be tracked using pre–post evaluations, and the associated outcomes should inform adaptive retraining strategies.

(2) Develop and Implement Standardized Guidelines: Introducing detailed and operational data entry guidelines will significantly improve the consistency and clarity of work order records. These guidelines should include (a) standardized templates for high-frequency issues, (b) a reference dictionary of recommended expressions and prohibited colloquialisms, (c) annotated examples highlighting common recording mistakes and suggested corrections, and (d) clear formatting rules for dates, locations, and administrative terms. To ensure their widespread adoption, these guidelines should be embedded within the hotline platform and updated periodically based on data audit findings and user feedback.

(3) Leverage Intelligent Assistance Tools: Modern technologies such as natural language processing (NLP) and address verification systems can enhance the accuracy of recorded data. Intelligent tools should be integrated into the data entry interface to (a) provide **real-time alerts** for spelling errors and improper keywords, (b) offer **automatic suggestions** for standard issue descriptions based on historical data, (c) enable the use of APIs (e.g., Baidu Maps) to verify and complete address fields, and (d) automatically flag incomplete or semantically inconsistent records.

These tools can operate in a human-in-the-loop manner, where suggestions facilitate rather than replace staff decisions, ensuring both the efficiency and integrity of data.

(4) Implement Automated and Dual Review Mechanisms: Establishing a layered quality review system can further enhance the accuracy of data. The first layer involves automated scripts that detect missing values, format inconsistencies, and low-quality text. The second layer consists of a manual review of flagged records by supervisors. The results obtained through this process should feed into a monthly feedback and correction loop, providing targeted guidance to staff and informing future training needs. In the long term, this mechanism should be supported by dashboard analytics, allowing for the tracking of quality trends and the benchmarking of performance.

Through the adoption of these strategies, government departments can significantly improve the standardization and reliability of hotline data. These recommendations not only address existing shortcomings in data recording but also provide a clear path for continuous improvement, thereby optimizing workflow efficiency, enhancing service delivery, and increasing public satisfaction.

6. Limitations and Future Work

This study proposed a multi-dimensional scoring method for hotline data record standards and validated its effectiveness through empirical analysis. However, several limitations remain. First, the data used were limited to a single municipal hotline system, which may restrict the generalizability of the findings. Second, while the six considered indicators were assigned equal weights in the scoring model, the random forest results revealed varying levels of predictive importance among these indicators. Future work could explore dynamic or learning-based weighting schemes to improve the accuracy of associated methods. Additionally, the existing evaluation framework is based on histor-

ical static data and, as such, real-time monitoring and feedback mechanisms should be incorporated in future studies to support adaptive improvements in hotline service quality.

Author Contributions: Conceptualization, Z.Z. and T.Z.; methodology, Z.Z.; software, Z.Z.; validation, Z.Z. and T.Z.; formal analysis, Z.Z.; investigation, Z.Z.; resources, Z.Z.; data curation, Z.Z.; writing—original draft preparation, Z.Z.; writing—review and editing, T.Z.; visualization, Z.Z.; supervi-sion, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Research Start-up Foundation of Recruiting Talents of Nanjing University of Posts and Telecommunications, (grant number NY223210).

Informed Consent Statement: This article does not contain any studies with human participants or animals performed by any of the authors.

Data Availability Statement: Data Access: https://pan.baidu.com/s/1mstmbJodyT-j6XkSlrA8vw?pwd=sr12, accessed on 24 February 2025.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhu, X.; Chen, Z.; Cheng, T.; Yang, C.; Wu, D.; Wu, Y.; Wang, H. Blockchain for urban governance: Enhancing trust in smart city systems with advanced techniques. *Sustain. Cities Soc.* **2024**, *108*, 105438. [CrossRef]
2. Zhang, Z.C.; Li, A.G.; Wang, L.; Cao, W.; Yang, J. Big data-assisted urban governance: A comprehensive system for business documents classification of the government hotline. *Eng. Appl. Artif. Intell.* **2024**, *132*, 107997. [CrossRef]
3. The State Council of the People's Republic of China. Guidelines on Further Improving the 12345 Government Hotline Service Mechanism. 2021. Available online: https://www.gov.cn/zhengce/content/2021-01/06/content_5577419.htm (accessed on 6 January 2021).
4. Hunan Provincial People's Government General Office. Notice on Issuing the 2025 Key Points of Government Services and Data Work in Hunan Province. 2024. Available online: https://hunan.gov.cn/hnszf/xxgk/wjk/szfbgt/202406/t20240613_33326346.html (accessed on 31 May 2024).
5. Zheng, Y.; Gan, Q.; Zhang, C.; Zhang, X. The Current Situation and Challenges of Local Government Data Governance: An Empirical Study Based on 43 Government Hotline Departments. *E-Government* **2020**, *7*, 66–79.
6. Fan, B.; Yu, Y. Influencing Factors of Government Data Quality Based on Adaptive Structuration Theory: A Case Study of the Data from Government Service Hotline 12345. *Doc. Inf. Knowl.* **2021**, *38*, 13–24.
7. Wang, Y.Z.; Lian, D.F.; Chen, E.H. Event assignment based on KBQA for government service hotlines. *Appl. Artif. Intell.* **2024**, *38*, 2348162. [CrossRef]
8. Li, Z.; Yi, X.; Chen, S.; Zhang, J.; Li, T. Government Event Dispatch Approach Based on Deep Multi-view Network. *Comput. Sci.* **2024**, *51*, 216–222.
9. Zhang, Z. Understanding the relationship between normative records of appeals and government hotline order dispatching: A data analysis method. *Data Technol. Appl.* **2024**, *58*, 496–516. [CrossRef]
10. Huang, W.; Su, C.; Wang, Y. An intelligent work order classification model for government service based on multi-label neural network. *Comput. Commun.* **2021**, *172*, 19–24. [CrossRef]
11. Zhao, T.; He, Z.; Shao, J.; Regmi, A.; Shi, L.; Cai, Y. Decoding hotline's information with text mining: A protocol for improving tobacco control in Shanghai. *Tob. Induc. Dis.* **2024**, *22*, 107. [CrossRef]
12. Zhang, Z.; Li, A.; Xu, Y.; Liang, Y.; Jin, X.; Wu, S. Understanding citizens' satisfaction with the government response during the COVID-19 pandemic in China: Comprehensive analysis of the government hotline. *Libr. Hi Tech* **2023**, *41*, 91–107. [CrossRef]
13. Karkošková, S. Data governance model to enhance data quality in financial institutions. *Inf. Syst. Manag.* **2023**, *40*, 90–110. [CrossRef]
14. de Oliveira, N.R.; Medeiros, D.S.; Moraes, I.M.; Andreonni, M.; Mattos, D.M. Towards intent-based management for Open Radio Access Networks: An agile framework for detecting service-level agreement conflicts. *Ann. Telecommun.* **2024**, *79*, 693–706. [CrossRef]
15. Zhang, X.; Yu, X. Measurement and improvement of island B&BS customer satisfaction. *Acta Psychol.* **2024**, *248*, 104425.
16. Ali, O.; Shrestha, A.; Chatfield, A.; Murray, P. Assessing information security risks in the cloud: A case study of Australian local government authorities. *Gov. Inf. Q.* **2020**, *37*, 101419. [CrossRef]
17. Baffour, B.; Valente, P. An evaluation of census quality. *Stat. J. IAOS* **2012**, *28*, 121–135. [CrossRef]

18. Zhang, Z.; Lin, X.; Shan, S. Big data-assisted urban governance: An intelligent real-time monitoring and early warning system for public opinion in government hotline. *Future Gener. Comput. Syst.* **2023**, *144*, 90–104. [CrossRef]
19. Tejedo-Romero, F.; Ferraz Esteves Araujo, J.F.; Gonçalves Ribeiro, M.J. The usability of Brazilian government open data portals: Ensuring data quality. *Humanit. Soc. Sci. Commun.* **2025**, *12*, 297. [CrossRef]
20. Mohmad, N.; Hassan, N.; Samy, G.; Aziz, N.S.A.; Maarop, N.; Abu Bakar, N.A. A review of factors influencing patient readmission based on data quality dimension model. In Proceedings of the 2020 8th International Conference on Information Technology and Multimedia (ICIMU), Selangor, Malaysia, 24–25 August 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 26–31.
21. Park, J.; Tosaka, Y. Metadata quality control in digital repositories and collections: Criteria, semantics, and mechanisms. *Cat. Classif. Q.* **2010**, *48*, 696–715. [CrossRef]
22. Schmidt, C.; Struckmann, S.; Enzenbach, C.; Reineke, A.; Stausberg, J.; Damerow, S.; Huebner, M.; Schmidt, B.; Sauerbrei, W.; Richter, A. Facilitating harmonized data quality assessments: A data quality framework for observational health research data collections with software implementations in R. *BMC Med. Res. Methodol.* **2021**, *21*, 63. [CrossRef]
23. Song, H.; Yu, J.; Han, Q. Industrial multivariate time series data quality assessment method. *J. Comput. Appl.* **2024**, *44*, 1743–1750.
24. Vetro, A.; Canova, L.; Torchiano, M.; Minotas, C.O.; Iemma, R.; Morando, F. Open data quality measurement framework: Definition and application to open government data. *Gov. Inf. Q.* **2016**, *33*, 325–337. [CrossRef]
25. Ba, W.; Chen, B.; Li, Q. Comprehensive evaluation method for traffic flow data quality based on grey correlation analysis and particle swarm optimization. *J. Syst. Sci. Syst. Eng.* **2024**, *33*, 106–128. [CrossRef]
26. Wu, D.; Xu, H.; Wang, Y.; Zhu, H. Quality of government health data in COVID-19: Definition and testing of an open government health data quality evaluation framework. *Libr. Hi Tech* **2021**, *40*, 516–534. [CrossRef]
27. Deng, Y.; Ju, H.; Zhong, G.; Li, A.; Ding, Y. A general data quality evaluation framework for dynamic response monitoring of long-span bridges. *Mech. Syst. Signal Process.* **2023**, *200*, 110514. [CrossRef]
28. Qaiser, S.; Ali, R. Text mining: Use of TF-IDF to examine the relevance of words to documents. *Int. J. Comput. Appl.* **2018**, *181*, 25–29. [CrossRef]
29. Guo, Y.; Barnes, S.; Jia, Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. *Tour. Manag.* **2017**, *59*, 467–483. [CrossRef]
30. Tang, J.; Meng, Z.; Nguyen, X.; Mei, Q.; Zhang, M. Understanding the limiting factors of topic modeling via posterior contraction analysis. In Proceedings of the International Conference on Machine Learning, PMLR, Beijing, China, 21–26 June 2014; pp. 190–198.
31. Hadi, M.; Fard, F. AOBTM: Adaptive online biterm topic modeling for version-sensitive short-texts analysis. In Proceedings of the 2020 IEEE International Conference on Software Maintenance and Evolution (ICSME), Adelaide, Australia, 28 September–2 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 593–604.
32. Sousa, M.; Sakiyama, K.; Rodrigues, L.; Moraes, P.H.; Fernandes, E.R.; Matsubara, E.T. BERT for stock market sentiment analysis. In Proceedings of the 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), Portland, OR, USA, 4–6 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1597–1601.
33. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794.
34. Rau, L. Extracting company names from text. In Proceedings of the Seventh IEEE Conference on Artificial Intelligence Application, Miami Beach, FL, USA, 24–28 February 1991; IEEE Computer Society: Washington, DC, USA, 1991; pp. 29–32.
35. Chinchor, N. MUC-6 named entity task definition (version 2.1). In Proceedings of the 6th Message Understanding Conference, Columbia, MD, USA, 6–8 November 1995.
36. Fukuda, K.; Tamura, A.; Tsunoda, T.; Takagi, T. Toward information extraction: Identifying protein names from biological papers. In *Pacific Symposium on Biocomputing*; World Scientific Publishing Co. Inc.: Singapore, 1998; pp. 707–718.
37. Quimbaya, A.; Múnera, A.; Rivera, R.; Rodríguez, J.C.D.; Velandia, O.M.M.; Peña, A.A.G.; Labbé, C. Named entity recognition over electronic health records through a combined dictionary-based approach. *Procedia Comput. Sci.* **2016**, *100*, 55–61. [CrossRef]
38. Keretna, S.; Lim, C.; Creighton, D. A hybrid model for named entity recognition using unstructured medical text. In Proceedings of the 2014 9th International Conference on System of Systems Engineering (SOSE), Glenelg, Australia, 9–13 June 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 85–90.
39. Yamada, H.; Kudo, T.; Matsumoto, Y. Japanese named entity extraction using support vector machine. *Trans. Inf. Process. Soc. Jpn.* **2002**, *43*, 44–53.
40. Han, X.; Kwoh, C.; Kim, J. Clustering-based active learning for biomedical named entity recognition. In Proceedings of the International Joint Conference on Neural Networks, Vancouver, BC, Canada, 24–29 July 2016; IEEE: Piscataway, NJ, USA, 2016.
41. Li, J.; Sun, A.; Han, J.; Li, C. A survey on deep learning for named entity recognition. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 50–70. [CrossRef]
42. Ningrinla, M.; Rakesh, T. KNN-ST: Exploiting spatio-temporal correlation for missing data inference in environmental crowd sensing. *IEEE Sens. J.* **2021**, *21*, 3429–3436.

43. Rocchio, J. Relevance feedback in information retrieval. In *The Smart Retrieval System Experiments in Automatic Document Processing*; Prentice Hall: New York, NY, USA, 1971; pp. 313–323.
44. Li, Q.; Wan, P.; Hao, T. Application of PSO-SVM-based text classification on the precision marketing of insurance. *J. Glob. Trade Manag.* **2022**, *28*, 111–129.
45. Pipino, L.L.; Lee, Y.W.; Wang, R.Y. Data quality assessment. *Commun. ACM* **2002**, *45*, 211–218. [CrossRef]
46. Lewis, A.; Weiskopf, N.; Abrams, Z.; Foraker, R.; Lai, A.M.; O Payne, P.R.; Gupta, A. Electronic health record data quality assessment and tools: A systematic review. *J. Am. Med. Inform. Assoc.* **2023**, *30*, 1730–1740. [CrossRef]
47. Varela, L.; Sandhu, N.; Walker, R.; Southern, D.A.M.; Quan, H.; Eastwood, C.A.R. Development of data quality indicators for improving hospital international classification of diseases-coded health data quality globally. *Med. Care* **2024**, *62*, 575–582. [CrossRef]
48. Cai, L.; Zhu, Y. The challenges of data quality and data quality assessment in the big data era. *Data Sci. J.* **2015**, *14*, 2. [CrossRef]
49. Reda, O.; Benabdellah, N.C.; Zellou, A. A systematic literature review on data quality assessment. *Bull. Electr. Eng. Inform.* **2023**, *12*, 3736–3757. [CrossRef]
50. Chen, H.; Hailey, D.; Wang, N.; Yu, P. A review of data quality assessment methods for public health information systems. *Int. J. Environ. Res. Public Health* **2014**, *11*, 5170–5207. [CrossRef]
51. Zhou, Y.; Tu, F.; Sha, K.; Ding, J.; Chen, H. A survey on data quality dimensions and tools for machine learning. *arXiv* **2024**, arXiv:2406.19614.
52. Batini, C.; Cappiello, C.; Francalanci, C.; Maurino, A. Methodologies for data quality assessment and improvement. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 1–52. [CrossRef]
53. Wang, R.Y.; Strong, D.M. Beyond accuracy: What data quality means to data consumers. *J. Manag. Inf. Syst.* **1996**, *12*, 5–33. [CrossRef]
54. Hamdy, M.E.; Shaalan, K. A hybrid framework for applying semantic integration technologies to improve data quality. *Int. J. Inf. Technol.* **2018**, *2*, 27–44.
55. Tian, Z.; Zhong, R.Y.; Vatankhah Barenji, A.; Wang, Y.T.; Li, Z.; Rong, Y. A blockchain-based evaluation approach for customer delivery satisfaction in sustainable urban logistics. *Int. J. Prod. Res.* **2021**, *59*, 2229–2249. [CrossRef]
56. Redman, T.C. The impact of poor data quality on the typical enterprise. *Commun. ACM* **1998**, *41*, 79–82. [CrossRef]
57. Gharehchopogh, F.S.; Khalifelu, Z.A. Analysis and evaluation of unstructured data: Text mining versus natural language processing. In Proceedings of the 2011 5th International Conference on Application of Information and Communication Technologies (AICT), Azerbaijan, Baku, 12–14 October 2011; IEEE: Piscataway, NJ, USA; pp. 1–4.
58. Li, Y.; Shang, H. Service quality, perceived value, and citizens' continuous-use intention regarding e-government: Empirical evidence from China. *Inf. Manag.* **2020**, *57*, 103197. [CrossRef]
59. Cheng, J.; Liu, Y. The effects of public attention on the environmental performance of high-polluting firms: Based on big data from web search in China. *J. Clean. Prod.* **2018**, *186*, 335–341. [CrossRef]
60. Tao, F.; Qi, Q.; Liu, A.; Kusiak, A. Data-driven smart manufacturing. *J. Manuf. Syst.* **2018**, *48*, 157–169. [CrossRef]
61. Adegoke, K.; Adegoke, A.; Dawodu, D.; Bayowa, A.; Adekoya, A. Interoperability in Digital Healthcare: Enhancing Consumer Health and Transforming Care Systems. *Preprint* **2025**. [CrossRef]
62. Otto, B.; Weber, K. Data governance. In *Daten-und Informationsqualität: Auf dem weg zur Information Excellence*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 277–295.
63. Mutinda, F.W.; Yada, S.; Wakamiya, S.; Aramaki, E. Semantic textual similarity in Japanese clinical domain texts using BERT. *Methods Inf. Med.* **2021**, *60*, e56–e64. [CrossRef]
64. Huang, Y. Empirical study on the factors affecting the performance of domestic government data openness. *Soft Sci.* **2023**, *37*, 10–17.
65. Niu, X.; Wang, X.; Gao, J.; Wang, X. Has third-party monitoring improved environmental data quality? An analysis of air pollution data in China. *J. Environ. Manag.* **2020**, *253*, 109698. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Spillover Effects of Physicians' Prosocial Behavior: The Role of Knowledge Sharing in Enhancing Paid Consultations Across Healthcare Networks

Yuting Zhang and Jiantong Zhang *

School of Economics and Management, Tongji University, Shanghai 200092, China; z_y_t2019@tongji.edu.cn
* Correspondence: zhangjiantong@tongji.edu.cn

Abstract: This study investigates the spillover effects of physicians' prosocial behavior, specifically knowledge sharing, on the paid consultations of other physicians within the same specialty and offline hospital. Using data from an online healthcare platform, we apply propensity score matching to explore how the sharing of medical knowledge by physicians influences the consultation outcomes of their colleagues. The results reveal significant positive spillover effects, indicating that prosocial behavior benefits other physicians within the same specialty and healthcare institution, thereby enhancing collaboration within the healthcare ecosystem. The spillover effect is stronger within the same offline hospital's physicians on the online healthcare platform, suggesting that knowledge sharing has a more localized impact within the same healthcare institution. Furthermore, the study examines heterogeneity across both physician-level characteristics (e.g., popularity, title, price, gender) and contextual factors (e.g., specialty type, hospital level, wait time, regional GDP). The findings show that the magnitude and direction of spillover effects differ by subgroup, shaped by professional visibility, authority, and organizational structure. These insights contribute to the understanding of how prosocial behavior can foster collaboration and benefit healthcare networks beyond individual physicians, offering practical implications for healthcare platforms, administrators, and policymakers.

Keywords: online healthcare platform; spillover effect; prosocial behavior; knowledge sharing; paid consultation

1. Introduction

Online healthcare platforms (OHPs) have developed rapidly in recent years. They pool medical resources from different hospitals at various levels, offering patients the opportunity to access high-level doctors nationwide [1]. The patients can easily consult their desired doctors in terms of a series of paid services, including text-based, video, and voice consultations [2]. Normally, the consultation fee ranges between 10 and 300 yuan per consultation. Aside from the paid services, free services that OHPs provide include knowledge sharing [3]. Such services enable patients to research before deciding on their health. Usually, a patient would go through a doctor's profile before visiting them; this includes the expertise of such doctors, the ranking of the hospital in that particular field of medicine, and the doctor's specialty [4].

Knowledge sharing has been one of the most prominent services featured on physicians' profiles and has also been widely discussed in the literature on OHPs. It is typically offered by physicians, nursing staff, and various other healthcare practitioners who are dedicated to the distribution of medical information [3]. The objective is to establish a

virtual community that enables patients to gain a clearer comprehension of their health issues, simultaneously fostering reciprocal learning among healthcare professionals [5]. The exchange of knowledge not only bolsters patient trust but also facilitates informed decision-making concerning healthcare among patients [6]. This boosts engagement on the platform and raises the number of consultations [7]. The literature has already extensively discussed the importance of knowledge sharing in attracting patients and impacting their consulting behavior.

Knowledge sharing is mostly treated as a form of prosocial behavior [8,9]. Prosocial behavior involves both altruistic and egoistic motivations [10]. Altruism refers to acts that mostly benefit other people, while egoism involves acts that benefit the actor and other people [11]. Prior research indicates that knowledge sharing between doctors and patients fosters mutual trust, helps build credibility, and raises the professional profile. All these positive outcomes eventually attract more patients and bring improved financial returns [3,12]. On the other hand, pros for patients in an altruistic sense encompass a better understanding of one's health condition through medical knowledge gained, empowerment to make informed choices, and lessened asymmetry of information, which prevails between the caregiver and the patients [13]. While previous research has largely focused on the altruistic dimensions of knowledge sharing with patients, there has been a relative neglect of the ways in which physicians' prosocial behaviors may impact their colleagues within the same medical specialty or institution. This paper tries to fill this gap by bringing to the forefront the altruistic aspects of knowledge sharing and exploring its subsequent effects on other physicians.

Previous spillover effects studies have mainly addressed the domain of consumer behavior, where experiences with one product influenced decisions related to other products within the same brand [14]. Spillover effects, when it comes to competitors, typically explored in consumer behavior, have been mainly focused on the negative impacts, such as how a crisis can affect the financial value of competitors [15,16]. However, in the case of healthcare, spillover effects can extend beyond these negative impacts, potentially fostering a more cooperative environment. In the context of OHPs, the spillover effect goes further than the individual consultations of the doctor; it refers to the potential for one doctor's knowledge sharing to spill over into the consultations of other doctors within the same specialty or hospital. This phenomenon has not been fully explored in OHPs, especially in prosocial behavior. By investigating the spillover effects of knowledge sharing, this paper extends the scope of prosocial altruism from individual physicians to their colleagues in the broader healthcare ecosystem. Given this background, the current study attempts to answer the following fundamental research questions:

Q1: Does a doctor's knowledge sharing increase the number of paid consultations for other doctors in the same specialty on OHP?

Q2: Does a doctor's knowledge sharing increase the number of paid consultations for other doctors in the same offline hospital?

In addition, this study examines how the spillover effects of physicians' knowledge sharing are conditioned by a range of individual-level and contextual factors. At the individual level, physician characteristics such as online popularity, consultation price, professional title, and gender may influence how patients perceive and process information, thereby moderating the extent to which knowledge sharing leads to spillover effects. At the contextual level, institutional and environmental attributes—including specialty type, hospital level, consultation responsiveness (as proxied by wait time), and the economic development level of the hospital's province—may also shape the diffusion of trust and information. These dimensions reflect both the personal traits of the physician and the broader structural conditions under which medical knowledge is shared and interpreted.

We therefore explore the following research questions:

Q3: How are these spillover effects moderated by physician-level characteristics, such as popularity, consultation price, professional title, and gender?

Q4: How do contextual and institutional factors—including specialty type, hospital level, consultation responsiveness, and the regional economic environment—condition the magnitude of the spillover effects?

This paper empirically investigates the spillover effects of physicians' prosocial behavior, specifically knowledge sharing, on online healthcare platforms. Our findings indicate that, besides helping patients, such prosocial behavior exerts a strong influence on the paid consultations of other physicians within the same specialty and offline hospital. More importantly, the magnitude of these spillover effects varies systematically across both physician-level characteristics and contextual factors. Stronger spillover effects are observed among physicians in tertiary hospitals and in specialties such as obstetrics and gynecology, as well as among those with distinct individual traits such as high reputation or junior title—reflecting the complex interplay between organizational position, individual visibility, and contextual receptiveness. These findings highlight the importance of knowledge sharing as a prosocial behavior that can enhance cooperation and mutual learning among healthcare professionals, leading to improved patient outcomes and increased platform engagement.

This study contributes to theory in several ways. First, it extends our understanding of prosocial behavior by examining its indirect effects on the healthcare ecosystem. While previous studies have emphasized the direct benefits of prosocial behavior for either patients or individual physicians, this research highlights its role in creating a reciprocal benefit for other physicians in the same network and thus fills an important gap in the current literature. Second, the study advances the theoretical understanding of spillover effect heterogeneity by jointly analyzing how both physician-level characteristics (e.g., popularity, price, title) and institutional factors (e.g., specialty, hospital level, responsiveness, regional GDP) moderate the strength and direction of these effects. The findings highlight the contextual features of prosocial behavior, which may have very different effects depending on both individual and situational factors. Third, the study contributes to understanding the interdependence between digital and physical healthcare systems, showing that online altruistic behavior can generate tangible benefits within offline hospital environments through increased trust diffusion and brand alignment.

2. Literature Review

2.1. Online Healthcare Platform and Knowledge Sharing

OHPs act as "medical associations", connecting major hospitals and grassroots medical institutions [17]. Through these platforms, patients gain the opportunity to consult with doctors nationwide from the comfort of their homes [18]. The platforms aggregate doctors from diverse healthcare institutions nationwide, offering a variety of services from top specialists to doctors at different levels and hospitals [19]. Paid services include text-based consultations, video consultations, and voice consultations, with prices typically ranging from 10 to 300 yuan per consultation. Additionally, they provide complimentary services such as knowledge sharing [20]. The diverse array of services attracts an increasing number of patients to utilize OHPs [21,22].

Knowledge sharing is a vital service for doctors to maintain a competitive edge [5]. Beyond merely enhancing doctors' reputations and increasing demand for their services, knowledge sharing also plays a pivotal role in providing patients with professional disease-related education, thereby benefiting their understanding and well-being [23–25]. Before consulting a doctor, patients often visit the relevant doctors' pages, where they may peruse

the complementarily offered knowledge sharing on their respective conditions, offered voluntarily by doctors [9].

Previous literature has widely investigated knowledge sharing in relation to digital health platforms. The benefits and costs framework based on social exchange theory has been used to explore the motivations for knowledge sharing [26]. It has also been found that knowledge sharing by doctors on digital health platforms can be positively influenced by the monetary rewards received from consultations [27]. Furthermore, the act of information sharing has been highlighted to have a significant effect on patient engagement and health outcomes [9]. It is also instrumental in attracting patients to utilize online health services, thus increasing the overall effectiveness of these platforms [28,29].

From the perspective of physicians, therefore, the impact of knowledge sharing on patients' decisions about consultations directly relates to their income. It is the relational aspect that brings to light the pragmatic effects of sharing knowledge but also resonates with the more general concept of prosocial behavior, which embraces both altruistic and self-serving motivations. Building on this, the next section deals with knowledge sharing as a form of prosocial behavior on OHPs.

2.2. Physicians' Prosocial Behavior

Prosocial behavior, which refers to actions taken with the altruistic motivation of helping others, spans a broad range of expressions, including helping strangers, engaging in volunteer activities, and making personal donations [30–33]. Research into prosocial behavior has tended to divide its motivations into two broad categories [11]. The key difference lies in whether the actor derives personal benefits, while the common feature shared by these categories is their essentially altruistic nature [30,31]. The initial classification characterizes prosocial behavior as entirely altruistic, whereby actions are exclusively motivated by the desire to assist others, devoid of any anticipation of personal advantage [34,35]. Conversely, the second classification adopts a more intricate viewpoint, wherein altruistic behaviors exist alongside self-serving intentions. From this perspective, individuals seek a mutually advantageous outcome, facilitating benefits for others while concurrently promoting their own interests [34,36–38]. The duality of motives is of vital importance in the context of OHPs.

In the present study, we adopt the latter perspective, according to which prosocial behavior is a construct that comprises both egoistic and altruistic motivations. In the OHP context, altruism has typically been related to the benefits that patients receive. For instance, prosocial behavior by doctors often enhances patients' understanding of their illness, strengthens their medical knowledge, and empowers them to make informed health-related choices. On the other hand, egoism refers to the personal benefits that physicians can derive from such actions, including increased reputation, greater patient trust, and, ultimately, higher consultation revenue. These altruistic and egoistic motivations often co-exist and drive physicians to perform prosocial behaviors on OHPs [39].

The impact of a doctor's prosocial behavior on their own consultations has been widely studied. Prior research has shown that doctors' prosocial behavior can increase patient trust, improve the doctor–patient relationship, and ultimately lead to more paid consultations for the physician sharing knowledge [3,26,40]. While altruism appears as a multifaceted construct in prosocial behavior, most studies under the OHP framework have focused primarily on the direct benefits of physicians' prosocial behavior for their patients [3,41].

Thus, a gap exists in understanding the consequences of physicians' prosocial behavior, viewed as altruistic behavior, for other physicians within the same healthcare network. This gap is addressed in the present study, which investigates the altruistic dimensions

of prosocial behavior in OHPs, focusing on the impact of physicians' prosocial behavior on colleagues. In so doing, we extend the understanding of prosocial behavior in digital healthcare platforms to include its effects beyond patients and include other healthcare providers within the network.

2.3. Spillover Effect

The spillover effect refers to the influence of one entity or action on related attributes, extending beyond its immediate impact by altering beliefs or behaviors about attributes not directly addressed, often through inferred associations [42]. The concept has been widely applied in many areas, including consumer behavior and organizational studies, in which the effects of one entity affect others in direct and indirect ways [43,44]. As an example, empirical research has shown that happiness can spill over to online prosocial behavior, hence influencing people's behavior in a broader social network [45]. The extant literature on spillover effects has so far concentrated on two major fields.

The first one, intra-brand spillover effects, examines how the behavior/attributes of one product influence consumer behavior about other products under the same brand [46]. For instance, a positive experience with a given product can enhance perceptions of other products from the same brand, thereby increasing the brand's overall equity [47,48]. The second area, inter-brand spillover effects, examines the impact that the behaviors of a specific product or brand have on its competitors [49]. For example, a creative new product development or marketing campaign may siphon off customers from competing products [50,51]. Nevertheless, most of the literature on inter-brand spillover effects is more focused on the negative implications, mainly studying how a brand crisis negatively impacts the financial market values of non-impacted competitors [52,53]. However, research into spillover effects among rivals that help in creating a collaborative environment is relatively limited.

The existing gap in academic research is particularly pertinent when looking at online healthcare, where the spillover effects may take on a more collaborative nature. Specifically, within the context of OHPs, the prosocial actions exhibited by physicians, such as knowledge sharing, may lead to the creation of a cooperative ecosystem that serves to benefit not only the individual physicians but also the broader platform environment [54,55]. Unlike the competitive dynamics typically observed in traditional consumer markets, knowledge sharing within OHPs may create spillover effects that go beyond individual gains and affect not only the physicians' practices but also those of other physicians within the same platform [56]. By examining the evolution of these dynamics, the current study expands the research canvas to include both digital and traditional healthcare systems, emphasizing the interrelated nature of healthcare ecosystems and the potential for knowledge diffusion to create synergistic benefits in both realms.

3. Research Hypothesis

This study examines the impact of physicians' altruistic behavior, specifically knowledge sharing, on the paid consultations of other physicians in the same specialty group or offline medical facility on online health platforms (OHPs). We focus on how physicians' prosocial behaviors, such as sharing medical knowledge, influence the paid consultations of other physicians in the same specialties or medical facilities. Furthermore, the extent and direction of these spillover effects may not be uniform. Drawing from the literature on digital reputation, healthcare decision-making, and institutional dynamics, we propose that both individual-level characteristics (e.g., popularity, price, professional title, and gender) and contextual factors (e.g., specialty type, hospital tier, consultation responsiveness, and

regional socioeconomic environment) moderate the strength of these effects. Figure 1 presents the conceptual framework guiding our hypotheses.

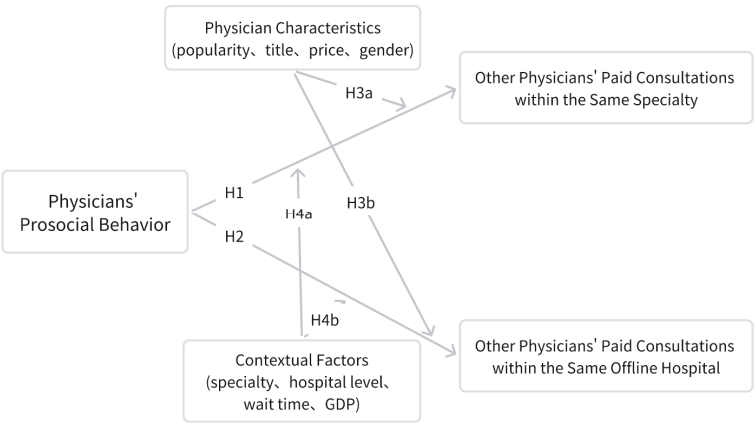


Figure 1. Conceptual Structure.

3.1. Impact of a Physician's Prosocial Behavior on Paid Consultations Within the Same Specialty

Previous studies have shown that physicians' prosocial behavior, such as sharing medical knowledge, results in more patient trust, which consequently increases consultations for the physician exhibiting this behavior. However, a physician's prosocial behavior may also have broader effects. In the case of OHPs, prosocial behavior may not only affect the individual physician but also spill over to affect consultations of other physicians in the same specialty. This occurs because those who come into contact with the collective knowledge of a certain doctor may develop a higher level of trust in the specialty in general, leading them to seek the consults of other doctors within the same field.

H1: *Physicians' prosocial behavior increases the number of paid consultations for other physicians within the same specialty group on OHP.*

3.2. Impact of a Physician's Prosocial Behavior on Paid Consultations Within the Same Offline Hospital

Similarly, knowledge sharing by physicians within an OHP can spill over to affect the consultations of physicians at the same offline hospital. Engaging a physician's shared knowledge may make the patients more trustful of the whole healthcare institution so as to make them visit other physicians also in the hospital for consultations. This spillover effect occurs as the patients' perceptions of the hospital's competence are influenced by the information given by the individual physicians, thus increasing consultations for other physicians within the same institution.

H2: *Physicians' prosocial behavior leads to increased paid consultations for other physicians within the same offline hospital on OHP.*

3.3. Moderating Effect of Physician Characteristics

Individual physician attributes may condition how knowledge-sharing behavior produces spillover effects. Characteristics such as online popularity, consultation price, professional title, and gender influence patient perceptions, information processing, and decision-making. These characteristics can either amplify the reach of knowledge-based trust, constrain its diffusion, or alter the underlying mechanisms through which trust is transferred. Accordingly, the spillover effects of prosocial behavior may vary across physicians within the same specialty group or within the same offline hospital, depending on these individual-level factors.

H3a: *The spillover effect of physicians' prosocial behavior within the same specialty group is moderated by physician-level characteristics (i.e., popularity, price, title, and gender).*

H3b: *The spillover effect of physicians' prosocial behavior within the same offline hospital is moderated by physician-level characteristics (i.e., popularity, price, title, and gender).*

3.4. Moderating Effect of Medical Context and Environment

Beyond individual attributes, contextual and institutional factors such as specialty type, hospital level, consultation responsiveness (proxied by wait time), and regional economic environment may influence how prosocial behavior spreads. These factors shape the structure of medical collaboration, visibility of information, patient expectations, and the perceived credibility of medical collectives.

H4a: *The spillover effect of physicians' prosocial behavior within the same specialty group is moderated by medical context (i.e., specialty type, hospital level, consultation responsiveness, and regional GDP).*

H4b: *The spillover effect of physicians' prosocial behavior within the same offline hospital is moderated by medical context (i.e., specialty type, hospital level, consultation responsiveness, and regional GDP).*

Based on these hypotheses, the present study aims to extend our understanding of the spillover effects of prosocial behavior in the context of OHPs. Specifically, we investigate not only whether physicians' prosocial behavior generates spillover effects across organizational boundaries—within the same specialty and offline hospital—but also how individual attributes and contextual factors condition the magnitude and direction of these effects. This layered hypothesis framework builds upon existing literature while reflecting the multifaceted heterogeneity inherent in platform-mediated medical interactions.

4. Data and Model

4.1. Data

This study relies on data made available by Haodf, one of the leading platforms in China. The dataset comprises 64,156 physicians who were active on the platform in 2020, spanning a wide range of specialties and offline hospital affiliations. The data are derived from physicians' digital profiles and patient-facing service records, including structured fields collected by the platform and user interaction metadata.

The dataset captures key physician characteristics—such as professional title, hospital level, gender, and responsiveness—as well as platform-based service attributes, including consultation price diversity, product diversity, and popularity. The core treatment variable is a binary indicator of whether a physician engaged in knowledge sharing, defined by exceeding a threshold number of medical articles posted on their profile. To measure peer effects, the outcome variables are constructed as leave-one-out means of paid consultations among peer groups defined by the same specialty, offline hospital, or R_department, respectively.

Additionally, a set of contextual and lagged control variables were constructed to mitigate confounding, including log-transformed provincial health facility capacity (ln_healthfaci_hosprov) and lagged mean and variance of popularity, response time, and service diversity within organizational peer groups. The full list of variable definitions used in the analysis is provided in Table 1.

Table 1. Definition of variables for main model.

Variable	Variable Type	Description
d_knowledge	Treatment	Dummy variable indicating whether a physician engaged in knowledge sharing (1 = Yes).
dm_paid	Outcome	Mean paid consultations of other physicians in the same specialty group (excluding focal physician).
hm_paid	Outcome	Mean paid consultations of other physicians in the same offline hospital (excluding focal physician).
title	Control	Physician's professional rank (1 = lowest; 4 = highest).
gender	Control	Dummy variable: 1 = Male, 0 = Female.
r_isgradeahospital	Control	Dummy variable for whether the physician is affiliated with a Grade A tertiary hospital.
responsetimely	Control	Physician's average service responsiveness (leave-one-out mean, lagged).
popularity	Control	Platform-calculated composite score reflecting physician's digital influence and service engagement.
pricediversity	Control	Number of distinct price points set by physician for different consultation products (range: 0–6).
productdiversity	Control	Number of distinct service types offered by physician (range: 0–8).
ln_healthfaci_hosprov	Control	Natural log of regional health facility density in the hospital's province.

4.2. Model

We apply propensity score matching (PSM) to estimate the causal effects of physicians' prosocial behavior on the paid consultations of other physicians in the same specialty group or offline hospital. PSM helps to mitigate selection bias by matching treated and control groups based on observable characteristics, ensuring a more accurate comparison.

In this study, we propose two models to test the spillover effects of knowledge sharing among physicians.

Model 1 investigates how a physician's knowledge-sharing influences the number of paid consultations of other physicians within the same specialty group. This model controls for a range of physician-level and institutional factors, including r.

$$\begin{aligned} \text{dm_paid}_i = {} & \alpha + \beta_1 \cdot \text{d_knowledge}_i + \beta_2 \cdot \text{gdp_hospitalprovince}_i + \beta_3 \cdot \text{title}_i + \beta_4 \cdot \text{gender}_i + \beta_5 \cdot \text{waittime}_i \\ & + \beta_6 \cdot \text{popularity_lm_w}_i + \beta_7 \cdot \text{popularity_lv_w}_i + \beta_8 \cdot \text{responsetimely_lm_w}_i + \beta_9 \cdot \text{responsetimely_w}_i \\ & + \beta_{10} \cdot \text{pricediversity_lm_w}_i + \beta_{11} \cdot \text{pricediversity_lv_w}_i + \beta_{12} \cdot \text{ln_healthfaci_hosprov_lm_w}_i \\ & + \beta_{13} \cdot \text{ln_healthfaci_hosprov_lv_w}_i + \beta_{14} \cdot \text{productdiversity_lm_w}_i + \beta_{15} \cdot \text{productdiversity_lv_w}_i + \varepsilon_i \end{aligned} \tag{1}$$

Model 2 extends this analysis by examining how knowledge sharing affects the paid consultations of other physicians within the same offline hospital. Similar to Model 1, it includes the same set of control variables.

$$\begin{aligned} \text{hm_paid}_i = {} & \alpha + \beta_1 \cdot \text{d_knowledge}_i + \beta_2 \cdot \text{gdp_hospitalprovince}_i + \beta_3 \cdot \text{title}_i + \beta_4 \cdot \text{gender}_i + \beta_5 \cdot \text{waittime}_i \\ & + \beta_6 \cdot \text{popularity_lm_w}_i + \beta_7 \cdot \text{popularity_lv_w}_i + \beta_8 \cdot \text{responsetimely_lm_w}_i + \beta_9 \cdot \text{responsetimely_lv_w}_i \\ & + \beta_{10} \cdot \text{pricediversity_lm_w}_i + \beta_{11} \cdot \text{pricediversity_lv_w}_i + \beta_{12} \cdot \text{ln_healthfaci_hosprov_lm_w}_i \\ & + \beta_{13} \cdot \text{ln_healthfaci_hosprov_lv_w}_i + \beta_{14} \cdot \text{productdiversity_lm_w}_i + \beta_{15} \cdot \text{productdiversity_lv_w}_i + \varepsilon_i \end{aligned} \tag{2}$$

By utilizing PSM, both models aim to provide a clearer understanding of how prosocial behavior, particularly knowledge sharing, generates spillover effects across healthcare settings.

5. Results

5.1. Analysis

In this study, we apply propensity score matching (PSM) to estimate the causal effects of physicians' prosocial behavior, specifically knowledge sharing, on the paid consultations of other physicians. PSM helps mitigate selection bias by matching treated and control groups based on observable characteristics, ensuring more accurate comparisons.

Spillover Effects Within Same Specialty Group and Same Offline Hospital (H1 and H2)

The analysis of spillover effects from knowledge sharing on paid consultations within the same specialty group and within the same offline hospital's physicians on the online healthcare platform (OHP) is presented in Table 2. We focus primarily on the matched sample results, as the unmatched samples may be confounded by unobserved factors, making them unsuitable for causal inference.

Table 2. Spillover effects of knowledge sharing on paid consultations within same specialty group and offline hospital (Average Treatment Effect on the Treated).

Dependent Variables	Sample	Treated	Controls	Difference	S.E.	T-Stat
dm_paid	Unmatched	923.039	911.627	11.411	3.654	3.120 ***
	Matched	923.038	897.378	25.661	4.252	6.040 ***
hm_paid	Unmatched	991.907	925.826	66.081	5.335	12.390 ***
	Matched	991.977	933.687	58.291	6.360	9.160 ***

Note (p): *** indicates significance at 1% level, respectively.

For the dependent variable dm_paid, the results show a positive and statistically significant spillover effect after matching. For the treated group, the Average Treatment Effect on the Treated (ATT) is 923.038, while the control group has 897.378, with a difference of 25.661 (t-stat = 6.040, ***), indicating a positive and statistically significant spillover effect. This finding supports Hypothesis 1, suggesting that knowledge sharing by physicians positively influences paid consultations within the same specialty group.

Similarly, for the dependent variable hm_paid, the matched sample results show a significant spillover effect. For the treated group, the ATT is 991.977, compared with 933.687 in the control group, with a difference of 58.291 (t-stat = 9.160, ***), demonstrating a positive and statistically significant spillover effect on paid consultations within the same offline hospital's physicians on OHP. This supports Hypothesis 2, confirming that knowledge sharing by physicians extends its effect to consultations among the same offline hospital physicians on OHP.

While the spillover effects observed in the unmatched sample are significant, they are not suitable for causal interpretation due to the potential for confounding factors. Therefore, the findings from the matched samples provide more reliable evidence, such as robustness checks, subgroup heterogeneity analysis, sensitivity tests, and placebo tests, for the hypotheses tested.

The spillover effect for hospital-level consultations (hm_paid, ATT = 58.291) is significantly larger than that for department-level consultations (dm_paid, ATT = 25.661), indicating that physicians' knowledge sharing has a broader and more pronounced influence on consultations among doctors within the same offline hospital compared with those within the same specialty group. This difference reflects the enhanced scope of

influence that arises from hospital-level resource integration and brand synergy, which enable platform complementarity and amplify the diffusion of trust and knowledge across the entire institutional network. Empirical evidence shows that hospitals with strong resource integration capabilities achieve substantially stronger spillover effects than those relying on isolated departmental efforts and that hospital branding further consolidates patient trust and increases the visibility of physicians across departments within the same institution [56].

5.2. Heterogeneity Tests

In this section, we examine the heterogeneity of the spillover effects of physicians' prosocial behavior across a range of physician-level and contextual-level dimensions. These include doctor popularity, consultation price, professional title, gender, medical specialty, hospital level, consultation wait time, and regional economic development. Table 3 summarizes the categorical variables used in the heterogeneity analysis, including their label names as shown in the results table and corresponding definitions or coding schemes.

Table 3. Variable descriptions and coding for heterogeneity analysis.

Variable	Description
popularity	Tertile based on physician's popularity index (1 = lowest, 3 = highest)
meanprice	Categorical grouping: Cheapest, Low, Mid, High, Premium
title	1 = Other doctors (e.g., resident or assistant physicians; lowest rank), 2 = Attending, 3 = Associate Chief, 4 = Chief Physician
gender	0 = Male, 1 = Female
specialty	Dummy variable for department groups, e.g., int_dum, obg_dum
hospitallevel	1 = Level 1 (lowest), 2 = Level 2, 3 = Level 3 (highest)
waittime	0 = None, 1 = Normal, 2 = Slow, 3 = Slower, 4 = Faster, 5 = Fast. Reflects average wait time as service efficiency.
gdp_hospitalprovince	Tertile of the GDP level of the hospital's province: Low, Medium, High.

In both Sections 5.2.1 and 5.2.2, we report the Average Treatment Effect on the Treated (ATT), which represents the average difference in paid consultation outcomes between physicians who engaged in knowledge sharing (treated group) and those who did not (control group) after propensity score matching. This measure isolates the effect of knowledge sharing by comparing each treated physician to statistically similar control physicians across subgroups. A positive and significant ATT indicates that knowledge sharing has a spillover effect on the matched peers' consultation volumes.

5.2.1. Spillover Effects Within the Same Department (dm_paid)

This subsection provides empirical tests for Hypotheses H3a and H4a, which explore how physician-level attributes and contextual medical environments moderate the spillover effects of prosocial behavior within the same specialty group. To investigate the heterogeneous spillover effects of knowledge sharing at the specialty level, we constructed subgroup variables based on the dimensions outlined in Section 5.2 and grounded in prior literature on peer effects, digital health platforms, and healthcare organizational behavior (see Table 4).

The results for Doctor Popularity show consistently significant and positive spillover effects across all groups (ATT = 28.71, 27.11, and 26.41 for low, medium, and high popularity, respectively; all t-values > 3, $p < 0.01$). Low-popularity physicians generated the strongest effect, indicating that their high-frequency knowledge-sharing helps reduce informational gaps and enhances collective trust in the department [21]. Medium-popularity physicians exhibited slightly weaker effects, likely due to the absence of both reputational authority

and algorithmic visibility—an issue known as the "moderate competitiveness trap" [57]. For high-popularity physicians, knowledge sharing tends to reinforce their personal brand rather than facilitate peer trust diffusion, producing a "halo effect" that limits spillover to colleagues [27].

Table 4. Heterogeneous treatment effects of knowledge sharing on dm_paid (2020, matched ATT results).

Group	Treated Mean	Control Mean	Difference ATT	Std. Error	t-Value
Doctor Popularity = Low	929.852	901.141	28.711	7.551	3.8 ***
Doctor Popularity = Medium	913.262	886.152	27.11	6.349	4.27 ***
Doctor Popularity = High	925.614	899.205	26.41	8.449	3.13 ***
Consultation Price = Cheapest	911.634	886.55	25.084	4.884	5.14 ***
Consultation Price = Low	957.155	945.441	11.715	14.279	0.82
Consultation Price = Mid	926.445	899.256	27.19	13.581	2 **
Consultation Price = High	968.145	946.055	22.09	24.026	0.92
Consultation Price = Premium	889.227	869.888	19.339	39.185	0.49
Physician Title = 1	969.395	926.165	43.23	22.026	1.96 **
Physician Title = 2	972.246	932.503	39.742	8.745	4.54 ***
Physician Title = 3	915.06	888.631	26.429	7.105	3.72 ***
Physician Title = 4	898.971	879.561	19.41	6.868	2.83 ***
Gender = Male	886.416	862.793	23.623	5.056	4.67 ***
Gender = Female	1012.209	990.881	21.328	7.202	2.96 ***
Specialty = Internal Medicine	661.02	649.781	11.24	4.014	2.8 ***
Specialty = TCM	1142.815	1122.394	20.421	9.657	2.11 **
Specialty = OB-GYN	964.09	947.076	17.014	7.331	2.32 **
Specialty = Pediatrics	1044.066	1014.549	29.517	15.937	1.85 *
Hospital Level = 1	951.531	962.995	−11.464	146.638	−0.08
Hospital Level = 2	948.748	936.97	11.778	20.529	0.57
Hospital Level = 3	921.779	894.307	27.472	4.376	6.28 ***
Wait Time = 0	868.166	848.069	20.096	5.697	3.53 ***
Wait Time = 1	954.522	938.315	16.207	16.58	0.98
Wait Time = 2	938.382	914.153	24.229	18.386	1.32
Wait Time = 3	935.092	896.261	38.83	20.915	1.86 *
Wait Time = 4	980.477	921.408	59.069	17.318	3.41 ***
Wait Time = 5	952.251	929.252	22.999	8.3	2.77 ***
Hospital GDP = Low	927.768	892.869	34.899	7.089	4.92 ***
Hospital GDP = Medium	926.579	893.175	33.404	7.781	4.29 ***
Hospital GDP = High	913.136	903.01	10.126	7.445	1.36

Note (p): *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

In terms of consultation price, physicians in the mid-price group displayed the highest spillover effect (ATT = 27.19, t = 2.00, **), followed by the lowest-price group (ATT = 25.08, t = 5.14, ***), while high-price segments showed no significant impact. This pattern suggests that within the same specialty, knowledge sharing by low-price physicians may reduce patients' decision-making risks, encouraging them to seek consultation from other doctors in the department. Mid-price physicians also benefit from informational spillover, although the significance is marginal. Prior research indicates that price-sensitive patients rely more on collective information to assess physician quality, with knowledge sharing expanding trust toward the entire department [58]. By contrast, patients in high-price groups tend to have concentrated preferences and loyalty, making them less susceptible to trust diffusion via knowledge sharing [27,40].

For the physician title, a clear gradient is observed. Junior physicians (Title = 1) exhibited the strongest spillover effect (ATT = 43.23, t = 1.96, **), followed by attending physicians (Title = 2, ATT = 39.74, t = 4.54, ***), associate chief physicians (Title = 3, ATT = 26.43, t = 3.72, ***), and chief physicians (Title = 4, ATT = 19.41, t = 2.83, ***). Although all groups show statistically significant positive effects, there is a general decline in ATT magnitude with higher professional ranks. This suggests that knowledge sharing by junior physicians plays an important role in reducing patient selection barriers [21]. Mid-level physicians often participate in collaborative activities such as case discussions and standard-setting, making their shared knowledge more actionable for peers [19]. In contrast, chief physicians may inadvertently centralize patient trust around themselves due to their authority, reducing the diffusion of trust within the department [27].

Regarding gender, both male (ATT = 23.62, t = 4.67, ***) and female (ATT = 21.33, t = 2.96, ***) physicians demonstrated significant positive spillover effects, with relatively small differences. This suggests that gender does not play a substantial role in moderating spillover effects within the same specialty.

The heterogeneity analysis across specialty types focused on the top five departments by consultation volume: obstetrics and gynecology, pediatrics, traditional Chinese medicine, internal medicine, and surgery. The surgical department was excluded from the analysis due to complete separation issues in the dataset, which prevented convergence of the propensity score matching (PSM) model. Following statistical best practices, departments with severe model violations—such as non-convergence caused by complete separation—were excluded to ensure validity. The analysis of specialty type reveals that pediatrics (ATT = 29.52, t = 1.85, *), traditional Chinese medicine (ATT = 20.42, t = 2.11, **), and obstetrics and gynecology (ATT = 17.01, t = 2.32 **) showed stronger spillover effects than Internal Medicine (ATT = 11.24, t = 2.8, ***). The high effect in pediatrics may stem from the fact that parents tend to rely on the department's collective expertise rather than on individual physician reputation, and knowledge sharing helps reduce informational asymmetry [40]. In traditional Chinese medicine, the long-term nature of chronic disease management encourages trust accumulation through continuous sharing [23]. For obstetrics and gynecology, knowledge sharing improves transparency and reduces patient concerns in high-sensitivity decisions [59]. Conversely, internal medicine covers a wide range of conditions, and information overload may dilute collective trust diffusion [23].

As for the hospital level, only Level 3 hospitals showed statistically significant spillover effects (ATT = 27.47, t = 6.28, ***), indicating that large, institutionalized hospitals have more structured pathways for horizontal knowledge diffusion. Patients in tertiary hospitals tend to trust the department as a whole, especially when lower-ranking doctors fill information gaps through knowledge sharing [21].

For wait time, which inversely reflects service efficiency, the results suggest a non-linear relationship. Physicians with extremely slow (Level 3, ATT = 38.83, t = 1.86, *) or extremely fast (Level 4: ATT = 59.07, t = 3.41, ***; Level 5: ATT = 22.99, t = 2.77, ***) response times produced significant spillover effects, as did those with no wait time (Level 0: ATT = 20.10, t = 3.53, ***). However, Levels 1 and 2 were not statistically significant. Extremely slow service may enhance trust through compensatory knowledge sharing [12], while fast responses increase patient satisfaction and visibility, forming "efficiency signals" that reinforce departmental credibility [2]. Moderate-efficiency doctors may fall into the "moderate competitiveness trap" [57].

Finally, under Regional GDP, significant spillover effects were observed in low (ATT = 34.90, t = 4.92, ***) and medium-GDP regions (ATT = 33.40, t = 4.29, ***), but not in high-GDP areas. In resource-constrained environments, patients tend to rely more on collective information and knowledge sharing to make informed decisions [21]. Medium-

GDP regions, while lacking absolute resource advantages, benefit more from departmental knowledge sharing due to their reliance on collective performance [60]. In contrast, patients in high-GDP regions may prioritize personal branding and physician credentials, limiting the impact of peer knowledge [27].

5.2.2. Spillover Effects Within the Same Hospital (hm_paid)

In alignment with Hypotheses H3b and H4b, this subsection examines whether the spillover effects of knowledge sharing within the same offline hospital vary across physician characteristics and institutional contexts. Using the same subgroup variables as in the specialty-level analysis, we analyze heterogeneous effects at the hospital level to enable direct comparison and deeper insight into how organizational structure and professional profiles shape the diffusion of shared knowledge (see Table 5).

Table 5. Heterogeneous treatment effects of knowledge sharing on hm_paid (2020, matched ATT results).

Group	Treated Mean	Control Mean	Difference ATT	Std. Error	t-Value
Doctor Popularity = Low	1119.404	1096.496	22.908	9.242	2.48 **
Doctor Popularity = Medium	922.558	913.281	9.276	8.849	1.05
Doctor Popularity = High	803.874	728.048	75.826	15.431	4.91 ***
Consultation Price = Cheapest	897.996	853.073	44.923	7.446	6.03 ***
Consultation Price = Low	1101.745	1125.443	−23.698	17.056	−1.39
Consultation Price = Mid	1167.833	1151.171	16.662	16.199	1.03
Consultation Price = High	1253.583	1261.245	−7.661	28.764	−0.27
Consultation Price = Premium	1350.66	1266.763	83.898	52.061	1.61
Physician Title = 1	953.874	919.559	34.315	30.48	1.13
Physician Title = 2	984.393	940.746	43.646	13.465	3.24 ***
Physician Title = 3	985.718	925.423	60.294	10.217	5.9 ***
Physician Title = 4	1005.139	929.345	75.794	10.018	7.57 ***
Gender = Male	966.497	916.793	49.703	7.635	6.51 ***
Gender = Female	1054.157	971.541	82.616	10.868	7.6 ***
Specialty = Internal Medicine	918.575	918.911	-0.336	13.689	−0.02
Specialty = TCM	1158.324	1141.465	16.859	33.272	0.51
Specialty = OB-GYN	1048.48	968.953	79.527	26.55	3 ***
Specialty = Pediatrics	1029.284	1065.853	−36.569	20.885	−1.75 *
Hospital Level = 1	1367.768	1418.339	−50.571	703.707	−0.07
Hospital Level = 2	782.639	664.682	117.957	78.641	1.5
Hospital Level = 3	997.65	942.901	54.749	6.01	9.11 ***
Wait Time = 0	927.019	873.115	53.903	9.446	5.71 ***
Wait Time = 1	1062.587	975.839	86.748	21.094	4.11 ***
Wait Time = 2	1030.328	957.875	72.453	24.572	2.95 ***
Wait Time = 3	1084.63	996.661	87.969	29.913	2.94 ***
Wait Time = 4	1077.371	1001.763	75.608	22.808	3.31 ***
Wait Time = 5	997.83	966.41	31.42	11.185	2.81 ***
Hospital GDP = Low	1011.917	977.831	34.087	10.375	3.29 ***
Hospital GDP = Medium	1048.373	990.537	57.836	11.376	5.08 ***
Hospital GDP = High	907.697	879.758	27.939	11.133	2.51 **

Note (p): *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

The analysis of physician popularity shows a non-linear pattern. High-popularity physicians generated the strongest spillover effect (ATT = 75.83, t = 4.91, ***), followed by low-popularity physicians (ATT = 22.91, t = 2.48 **), while the effect for medium-popularity

physicians was not statistically significant. This suggests the coexistence of two mechanisms: for high-popularity physicians, existing reputational capital—such as academic status or favorable patient reviews—acts as a "trust anchor". Their knowledge-sharing behavior signals hospital-wide competence, leading to trust transfer across colleagues [27]. In contrast, low-popularity physicians may use frequent knowledge sharing to fill information voids, thereby enhancing patient trust in the broader hospital network [21]. Medium-popularity physicians, lacking both reputational authority and algorithmic visibility, are more vulnerable to being overlooked, a phenomenon described as the "moderate competitiveness trap" [57].

For consultation price, the spillover effect was only significant in the lowest-price group (ATT = 44.92, t = 6.03, ***). Similar to the specialty-level findings, this suggests that price-sensitive patients are more likely to make use of publicly available information and exhibit greater responsiveness to knowledge-sharing cues [40]. Conversely, high-price consultations are typically associated with brand loyalty and patient inertia, weakening the collective trust diffusion effect [27,40].

Physician title exhibited a clear upward gradient in spillover strength. Chief physicians (Title = 4) showed the highest effect (ATT = 75.79, t = 7.57, ***), followed by associate chief physicians (Title = 3, ATT = 60.29, t = 5.90, ***), attending physicians (Title = 2, ATT = 43.65, t = 3.24, ***), while junior physicians (Title = 1) did not show a significant effect (ATT = 34.32, t = 1.13). These findings diverge from the specialty-level patterns, where junior physicians generated stronger spillovers. At the hospital level, knowledge sharing by senior doctors appears to drive broader influence, possibly due to their integration into the hospital's brand and increased algorithmic promotion. In particular, chief physicians are often publicly associated with the hospital's institutional image, amplifying the reach of their knowledge sharing and leveraging trust across departments [56,61]. Junior physicians, by contrast, may lack the visibility and authority to exert such influence, even when engaging in similar behaviors.

In the case of gender, female physicians demonstrated a stronger spillover effect (ATT = 82.62, t = 7.6, ***) than male physicians (ATT = 49.70, t = 6.51, ***). Prior studies suggest that female physicians may be more effective in emotional communication and cross-disciplinary coordination, which strengthens knowledge flows and facilitates trust-based collaboration within the hospital [60].

For specialty type, notable differences were observed. The surgical department was excluded from the hm_paid analysis, similar to its exclusion in the dm_paid analysis. Obstetrics and gynecology had the strongest spillover effect (ATT = 79.53, t = 3, ***), reflecting its high patient sensitivity and demand for information, which make it conducive to hospital-wide trust transfer via knowledge sharing [56]. In contrast, Pediatrics showed a negative spillover effect (ATT = −36.57, t = −1.75, *), potentially due to internal competition for limited resources, where high-frequency sharing from a few physicians attracts patients disproportionately, thereby reducing demand for others [62,63]. The effects for Traditional Chinese Medicine (ATT = 16.86, t = 0.51) and Internal Medicine (ATT = −0.336, t = −0.02) were not significant, possibly due to the lower relevance of hospital-level branding in these specialties or the dominance of offline, physician-specific care preferences [57,60].

With regard to hospital level, Level 3 hospitals again exhibited the strongest and statistically significant spillover effect (ATT = 54.75, t = 9.11, ***), consistent with the idea of platform-enabled "hospital–star physician" brand integration [56]. In these settings, patient trust tends to extend from high-reputation physicians to the institution itself, reinforcing horizontal knowledge diffusion.

For wait time, physicians with normal (Level 1, ATT = 86.75, t = 4.11, ***), slow (Level 2, ATT = 72.45, t = 2.95, ***), slower (Level 3, ATT = 87.97, t = 2.94, ***), and faster (Level 4,

ATT = 75.61, t = 3.31, ***) consultation response times exhibited the strongest spillover effects. These results reflect a trust synergy mechanism, whereby patients associate physicians with moderate efficiency profiles with the hospital's brand image, forming a "reliable and comprehensive" perception that enhances confidence in the entire institution [64]. In particular, physicians with slightly slower response times often compensate through sustained and in-depth knowledge sharing, which builds long-term trust among patients [65]. By contrast, physicians with no wait time (Level 0, ATT = 53.90, t = 5.71, ***) and the fastest response times (Level 5, ATT = 31.42, t = 2.81, ***) also demonstrated strong and significant spillover effects, though with slightly lower magnitudes. These physicians primarily benefit from efficiency signaling, which boosts algorithmic visibility and attracts more patients through perceived responsiveness. However, due to the limited depth of interaction, patient trust tends to remain at a surface level, which falls short of initiating the multistage process required for substantive trust development [66].

Finally, regional GDP did not significantly moderate the spillover effect at the hospital level—all three categories (low, medium, and high GDP) displayed significant positive effects (e.g., medium GDP: ATT = 57.84, t = 5.08, ***). This indicates that, compared with specialty-based effects, hospitals as higher-order organizational units are better equipped to transcend regional disparities. In particular, medium-GDP regions displayed the strongest spillover, consistent with theories of "latecomer advantage" in digital health infrastructure [40]. Meanwhile, in high-GDP areas, stronger physician networks and faster knowledge dissemination mechanisms help sustain significant spillover, albeit at a slightly lower magnitude [62].

In summary, knowledge sharing generates substantial spillover effects in both intra-specialty and intra-hospital contexts. However, the underlying mechanisms differ. At the specialty level, spillover is shaped by peer networks, trust reallocation among patients, and informational asymmetry. At the hospital level, platform visibility, brand alignment, and institutional routines jointly amplify knowledge diffusion, particularly when high-ranking physicians serve as brand anchors. These findings suggest that platforms should tailor knowledge-sharing incentives to account for organizational structure and physician heterogeneity in order to maximize the value of collective knowledge.

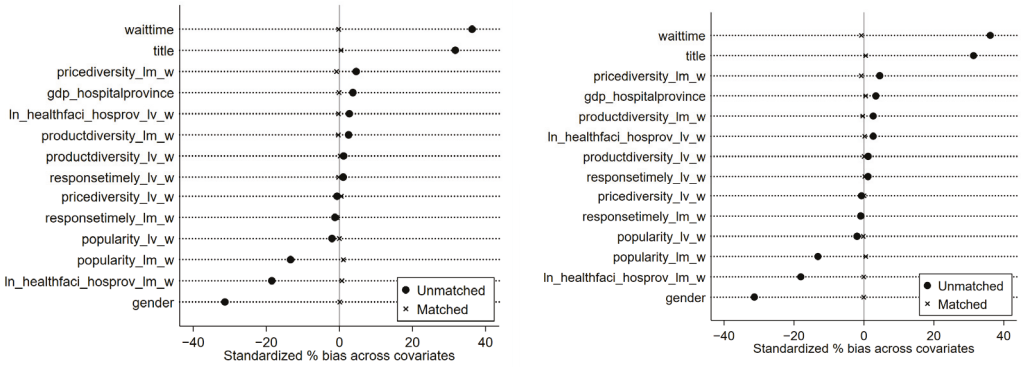
5.3. Matching Effectiveness and Balance Test

The *p*-test results shown in Figure 2a,b indicate the effectiveness of the matching process in minimizing bias across covariates. After performing the matching process, the standardized percentage bias for each covariate was reduced to below 10%, which is considered an acceptable threshold for achieving balance between the treated and control groups. This suggests that the matching process was successful in reducing bias, leading to a more reliable comparison between groups. The results for both dm_paid and hm_paid show that after matching, the bias in covariates was substantially minimized, confirming that the matching procedure has improved the comparability of the groups in the analysis. This demonstrates that the matching results are robust, and the subsequent analysis can be conducted with reduced confounding effects. Figure 2 displays the results for the main model.

5.4. Robustness Test

To ensure the reliability of the main findings, we conducted a series of robustness checks across multiple dimensions, including alternative matching parameters, resampled datasets, temporal subgroups, and estimation techniques. These robustness tests are designed to confirm the validity of the causal inferences underlying Hypotheses H1 and H2 and to ensure that the estimated treatment effects are not artifacts of model specification or

sampling variation. Specifically, the tests comprised caliper matching using a wider bandwidth (0.25 times the standard deviation), multistage sampling, year-specific subsamples, and inverse probability weighting (IPW) methods for both ATT and ATE estimation.



(**a**) (**b**)

Figure 2. Standardized mean difference graphs: (**a**) standardized mean difference graph of dm_paid; (**b**) standardized mean difference graph of hm_paid.

First, we re-estimated the models using a relaxed caliper of 0.25 standard deviations, as opposed to the default used in the main analysis [67]. The spillover effects of knowledge sharing on both dm_paid (within-specialty consultations) and hm_paid (within-hospital consultations) remained positive and statistically significant (t > 6.0), indicating that the results are not sensitive to the choice of matching threshold (see Table 6).

Table 6. Robust test (Average Treatment Effect on the Treated).

Outcome	Robustness Method	Treated	Controls	Difference	S.E.	t-Value
dm_paid	Caliper = 0.25 SD	923.038	897.374	25.665	4.252	6.04 ***
	Multistage Sampling	924.909	903.753	21.155	7.847	2.70 ***
	Subsample (Year 2018)	793.688	761.588	32.1	4.74	6.77 ***
	Subsample (Year 2019)	803.651	773.975	29.677	4.03	7.36 ***
hm_paid	Caliper = 0.25 SD	991.952	933.69	58.262	6.36	9.16 ***
	Multistage Sampling	990.802	923.964	66.838	11.244	5.94 ***
	Subsample (Year 2018)	835.572	790.576	44.996	6.317	7.12 ***
	Subsample (Year 2019)	850.717	795.035	55.681	5.454	10.21 ***

Note (p): *** indicates significance at the 1% level, respectively.

Second, using a multistage sampled dataset, the estimated ATT for dm_paid was 21.16 (t = 2.70 ***), while hm_paid reached 66.84 (t = 5.94 ***), reaffirming the consistency of the treatment effects even under substantial changes in sample structure (Table 6).

Year-specific robustness checks further support this pattern. For the dm_paid outcome, the ATT was 32.10 (t = 6.77 ***) in 2018 and 29.68 (t = 7.36 ***) in 2019. Correspondingly, for hm_paid, the ATT values were 44.99 (t = 7.12 ***) and 55.68 (t = 10.21 ***) in the respective years, showing temporal stability of the spillover effects across different annual cohorts (Table 6).

In addition, we implemented inverse probability weighting (IPW) to mitigate potential selection bias and offer an alternative identification strategy (see Table 7). Under IPW linear regression, the ATT estimate for dm_paid was 13.32 (t = 3.49 ***), and the ATE was 23.29 (t = 6.18 ***). For hm_paid, the corresponding ATT and ATE were 44.76 (t = 8.20 ***) and 61.64 (t = 11.18 ***), respectively. These results corroborate the causal interpretation of the main findings even after accounting for sample selection mechanisms.

Taken together, the spillover effects of physicians' knowledge sharing exhibit strong robustness across a wide range of methodological adjustments. The effects are consistently

more pronounced at the hospital level (hm_paid), underscoring the broader institutional and network-based transmission of informational externalities in platform-enabled healthcare environments.

Table 7. Robust Test Using Inverse Probability Weighting (IPW).

Outcome	Model	Effect	Coef.	S.E.	t-Value
dm_paid	IPW-ATT	ATT	13.32	3.81	3.49 ***
	IPW-ATE	ATE	23.29	3.77	6.18 ***
hm_paid	IPW-ATT	ATT	44.76	5.46	8.2 ***
	IPW-ATE	ATE	61.64	5.51	11.18 ***

Note (p): *** indicates significance at the 1% level, respectively.

5.5. Sensitivity Analysis

To test the robustness of the estimated treatment effects against potential unobserved confounders—such as physician motivation, managerial incentives, or patient selection—we conducted Rosenbaum bounds sensitivity analysis using binary outcomes defined as whether a physician's spillover effect exceeded the sample mean (i.e., dm_paid_high and hm_paid_high). For the specialty-level outcome (dm_paid_high), the effect remains statistically significant up to $\Gamma = 2.0$, implying that an unobserved confounder would need to more than double the odds of knowledge-sharing assignment to nullify the observed result. For the hospital-level outcome (hm_paid_high), the result holds up to $\Gamma = 1.3$, indicating that the hospital-level spillover effect is moderately sensitive to hidden bias. Overall, the analysis confirms that the findings are not easily attributable to unobserved heterogeneity and reinforces the credibility of the causal claims underlying Hypotheses H1 and H2 (see Table 8).

Table 8. Sensitivity analysis using Mantel–Haenszel bounds.

Variable	Gamma	Q_mh+	Q_mh−	p_mh+	p_mh−
dm_paid_high	1	2.084	2.084	0.0186	0.0186
	1.1	6.916	2.727	2.30×10^{-12}	0.0032
	1.2	11.331	7.138	0	4.70×10^{-13}
	1.3	15.397	11.2	0	0
	1.4	19.168	14.964	0	0
	1.5	22.685	18.474	0	0
	1.6	25.982	21.762	0	0
	1.7	29.086	24.856	0	0
	1.8	32.02	27.779	0	0
	1.9	34.801	30.549	0	0
	2	37.447	33.183	0	0
hm_paid_high	1	11.07	11.07	0	0
	1.1	6.232	15.915	2.30×10^{-10}	0
	1.2	1.818	20.347	0.0345	0
	1.3	2.222	24.431	0.0131	0
	1.4	5.982	28.222	1.10×10^{-9}	0
	1.5	9.483	31.76	0	0
	1.6	12.762	35.079	0	0
	1.7	15.844	38.204	0	0
	1.8	18.754	41.159	0	0
	1.9	21.51	43.962	0	0
	2	24.128	46.629	0	0

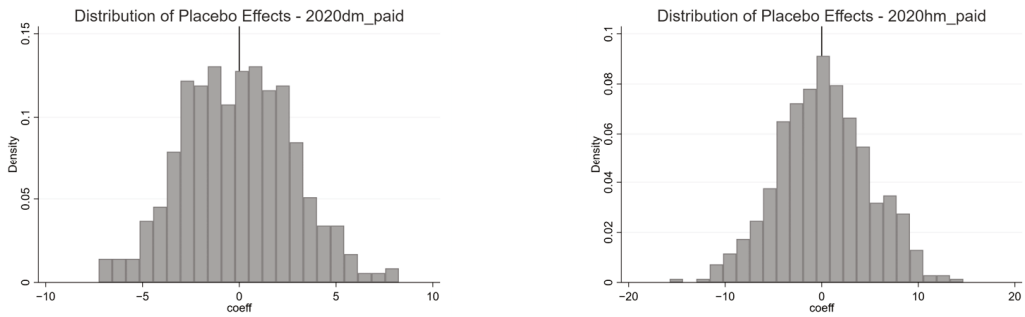
Gamma: odds of differential assignment due to unobserved factors; Q_mh+: MantelHaenszel statistic (assumption: overestimation of treatment effect); Q_mh−: Mantel–Haenszel statistic (assumption: underestimation of treatment effect); p_mh+: significance level (assumption: overestimation of treatment effect); p_mh−: significance level (assumption: underestimation of treatment effect).

5.6. Placebo Test

To assess the risk of false positives, we conducted placebo tests by randomly reassigning the treatment variable (d_knowledge) 500 times and re-estimating the treatment effect each time. As Table 9 shows, for the specialty-level outcome (dm_paid), the distribution of placebo coefficients is centered around zero (mean = −0.07), and only 3.8% of simulations yielded statistically significant results at the 5% level. For the hospital-level outcome (hm_paid), the placebo distribution also centers near zero (mean = 0.21), with a 4.8% false positive rate. These low rejection rates suggest that the observed spillover effects are unlikely to be driven by random chance, further strengthening the internal validity of the findings and reinforcing support for H1 and H2. Figure 3 displays the distribution of placebo effects.

Table 9. Placebo test—distribution of simulated treatment effects.

Outcome	True ATT	Mean Placebo Coeff	Std. Dev.	$p < 0.05$ (%)
dm_paid	25.66	−0.07	2.84	3.80%
hm_paid	58.29	0.21	4.76	4.80%



(**a**) (**b**)

Figure 3. Distribution of Placebo Effects (**a**) Distribution of Placebo Effects for dm_paid; (**b**) Distribution of Placebo Effects for hm_paid.

6. Discussion

6.1. Key Findings

This study investigates the spillover effects of physicians' prosocial behavior—specifically knowledge sharing—on the paid consultations of their peers within the same specialty group and the same offline hospital on an online healthcare platform (OHP). The findings provide strong empirical support for the presence of positive and statistically significant spillover effects, consistent with Hypotheses 1 and 2. Notably, the spillover effect is substantially larger within offline hospitals compared with within specialty groups. This difference suggests that the influence of knowledge sharing is more pronounced in hospital-level networks, where brand synergy and resource integration amplify patients' collective trust and facilitate broader diffusion of shared information across departments.

Beyond the confirmation of main effects, this study contributes a systematic investigation of spillover heterogeneity across two dimensions: (1) individual physician attributes (e.g., popularity, price, title, and gender) and (2) contextual/institutional characteristics (e.g., specialty type, hospital level, consultation responsiveness, and regional GDP). These multidimensional subgroup analyses reveal that the pathways and strength of spillover effects vary significantly across different contexts.

At the individual level, physician popularity moderates the direction and magnitude of spillover in distinct ways. Within specialty groups, less popular physicians exhibit stronger

spillovers due to their high-frequency sharing behavior that bridges information gaps and enhances departmental trust. In contrast, within hospitals, highly popular physicians produce stronger effects, likely due to their reputational capital acting as a trust anchor that diffuses confidence throughout the institutional brand. Physician title exhibits a reverse pattern: junior physicians generate larger spillovers within specialty groups, while senior physicians dominate at the hospital level. This divergence reflects how visibility, authority, and organizational anchoring condition the influence of knowledge-sharing behavior.

In terms of contextual heterogeneity, spillover effects are stronger in pediatrics, traditional Chinese medicine, and obstetrics and gynecology within specialty groups, highlighting the importance of collective trust and long-term engagement in patient decision-making. Conversely, at the hospital level, the strongest effect emerges in obstetrics and gynecology, while pediatrics exhibits a negative spillover—due to internal resource competition. Hospital level also matters: tertiary hospitals consistently show stronger spillovers, reaffirming the role of platform-enabled brand integration and horizontal collaboration. Additionally, while regional GDP moderates specialty-level spillovers, its influence is less evident at the hospital level, underscoring how institutional structures can mitigate regional disparities in healthcare access and trust formation.

Robustness checks, including alternative caliper widths, multistage sampling, year-specific subsamples, and inverse probability weighting (IPW) models, further support the validity of these findings. Across all methods, knowledge-sharing spillovers remain statistically significant and directionally consistent, with hospital-level effects consistently stronger. These robustness tests affirm the causal interpretation and reliability of the results across diverse empirical specifications. Moreover, the robustness of these findings is further supported by sensitivity and placebo analyses.

6.2. Theoretical Contributions

This study makes several important contributions to the literature. First, it extends the understanding of prosocial behavior by investigating its indirect effects within the ecosystem of healthcare. While prior work has focused almost exclusively on the direct benefits of prosocial behavior for patients or individual physicians [9,57], this study highlights its ability to generate reciprocal benefits for other physicians in the same network, thus filling a key gap in extant research.

The results contribute to the theoretical framework of spillover effect heterogeneity by distinguishing between physician-level and institutional-level moderators. Prior studies have often treated heterogeneity in isolation—focusing on single characteristics such as specialty or title—but this research adopts an integrative approach to jointly analyze how individual attributes and organizational context shape the strength, scope, and direction of spillover effects. The findings reveal that the same physician behavior may operate through different mechanisms depending on the surrounding institutional environment, offering a richer theoretical account of how prosocial actions propagate through complex systems.

This study contributes to the understanding of the interplay between online and traditional healthcare systems. It shows how online prosocial behavior can benefit hospital-wide visibility, trust coordination, and peer collaboration, thus underlining the interdependent nature of modern healthcare ecosystems.

6.3. Practical Implications

The findings offer actionable insights for OHPs, healthcare administrators, and policymakers. First, platforms can actively promote a culture of prosocial behavior by emphasizing its collective benefits. For platforms, promoting knowledge sharing can enhance not only individual physician visibility but also system-wide engagement and patient

trust. Importantly, incentive schemes should account for the observed heterogeneity: encouraging junior and less popular physicians to share content can strengthen specialty-level collaboration while leveraging the reputational capital of senior physicians enhances hospital-level diffusion.

Second, hospitals should integrate digital knowledge-sharing practices into their branding and trust-building strategies. Tertiary hospitals, in particular, stand to benefit from supporting high-profile physicians in content production, thereby reinforcing institutional credibility and patient engagement. Additionally, specialties such as obstetrics and gynecology, where spillovers are strong, may be prioritized in campaigns that promote digital outreach.

Lastly, policy frameworks should recognize prosocial behavior as a public good in healthcare ecosystems. By designing policies that reward information sharing and reduce structural barriers to trust diffusion, regulators can foster a more transparent, equitable, and patient-centered care environment.

6.4. Limitations and Future Directions

This study has certain limitations related to the availability of data from OHPs. While we have extended the research on spillover effects to include physicians within the same offline hospital, the inability to access additional offline hospital data for these patients limits the scope of our analysis. For example, data on patients' offline interactions with other physicians, their referral pathways, or their longitudinal healthcare engagement would provide a more comprehensive understanding of spillover dynamics across online and offline systems.

Future research could replicate this study across different online healthcare platforms or international contexts to test the robustness of the findings in varied institutional environments. Additionally, employing more advanced causal inference techniques—such as instrumental variable approaches or longitudinal panel designs—would allow for a deeper understanding of the long-term effects of prosocial behavior. Further exploration of heterogeneous forms of prosocial engagement, including peer recommendations, interactive Q&A, or multi-physician collaborations, could provide nuanced insights into trust dynamics and platform engagement mechanisms.

Author Contributions: Conceptualization: Y.Z. and J.Z.; Methodology: Y.Z. and J.Z.; Software: Y.Z.; Validation: Y.Z. and J.Z.; Formal Analysis: Y.Z.; Investigation: Y.Z.; Resources: Y.Z. and J.Z.; Data Curation: Y.Z.; Writing—Original Draft Preparation: Y.Z.; Writing—Review and Editing: Y.Z. and J.Z.; Visualization: Y.Z.; Supervision: J.Z.; Project Administration: Y.Z. and J.Z.; Funding Acquisition: J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Nos. 72371188, 71971156, 72271082, and 72071063), the Fundamental Research Funds for the Central Universities (No. 22120210241), and the Natural Science Foundation of Anhui Province (No. 2408085J041).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to thank the HaoDaiFu platform for providing the publicly available de-identified dataset used in this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ren, H.; Wang, L.; Wu, J. The faster or richer the response, the better performance? An empirical analysis of online healthcare platforms from a competitive perspective. *Decis. Support Syst.* **2024**, *184*, 114274. [CrossRef]
2. Liu, Y.; Zhang, X.; Liu, L.; Lai, K.H. Does voice matter? Investigating patient satisfaction on mobile health consultation. *Inf. Process. Manag.* **2023**, *60*, 103362. [CrossRef]
3. Zhang, X.; Liu, S.; Deng, Z.; Chen, X. Knowledge sharing motivations in online health communities: A comparative study of health professionals and normal users. *Comput. Hum. Behav.* **2017**, *75*, 797–810. [CrossRef]
4. Xiong, X.; Luo, L.; Zhou, S.; Li, V.J.; Zhou, Y.; Huo, Z. A profile of patients' and doctors' perceptions, acceptance, and utilization of e-health in a deprived region in southwestern China. *PLoS Digit. Health* **2023**, *2*, e0000238. [CrossRef]
5. Hara, N.; Foon Hew, K. Knowledge-sharing in an online community of health-care professionals. *Inf. Technol. People* **2007**, *20*, 235–261. [CrossRef]
6. Oh, S. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. *J. Am. Soc. Inf. Sci. Technol.* **2012**, *63*, 543–557. [CrossRef]
7. Han, G.; Kim, I. An Analysis of Factors Influencing Knowledge Sharing in Online Health Communities. *J. Inf. Syst.* **2024**, *33*, 51–74.
8. Kuvaas, B.; Buch, R.; Dysvik, A. Perceived training intensity and knowledge sharing: Sharing for intrinsic and prosocial reasons. *Hum. Resour. Manag.* **2012**, *51*, 167–187. [CrossRef]
9. Wang, J.J.; Liu, H.; Cui, X.; Ye, J.; Chen, H. Impact of a physician's prosocial behavior on the patient's choice: An empirical investigation in online health community. *Inf. Technol. People* **2023**, *36*, 1703–1725. [CrossRef]
10. Learning, S. Altruism and prosocial behavior. *Personal. Soc. Psychol.* **2003**, *5*, 463.
11. Pfattheicher, S.; Nielsen, Y.A.; Thielmann, I. Prosocial behavior and altruism: A review of concepts and definitions. *Curr. Opin. Psychol.* **2022**, *44*, 124–129. [CrossRef] [PubMed]
12. Singh, J.B.; Chandwani, R.; Kumar, M. Factors affecting Web 2.0 adoption: Exploring the knowledge sharing and knowledge seeking aspects in health care professionals. *J. Knowl. Manag.* **2018**, *22*, 21–43. [CrossRef]
13. Zhang, X.; Liu, S. Understanding relationship commitment and continuous knowledge sharing in online health communities: A social exchange perspective. *J. Knowl. Manag.* **2022**, *26*, 592–614. [CrossRef]
14. Erdem, T.; Sun, B. An empirical investigation of the spillover effects of advertising and sales promotions in umbrella branding. *J. Mark. Res.* **2002**, *39*, 408–420. [CrossRef]
15. Durnev, A.; Mangen, C. The spillover effects of MD&A disclosures for real investment: The role of industry competition. *J. Account. Econ.* **2020**, *70*, 101299.
16. Liu, D.; Varki, S. The spillover effect of product recalls on competitors' market value: The role of corporate product reliability. *J. Bus. Res.* **2021**, *137*, 452–463. [CrossRef]
17. Tan, H.; Zhang, X.; Yang, Y. Satisfaction or gratitude? Exploring the disparate effects of physicians' knowledge sharing on patients' service evaluation in online medical consultations. *Inf. Syst. J.* **2023**, *33*, 1186–1211. [CrossRef]
18. Ghosh, A.; Gupta, R.; Misra, A. Telemedicine for diabetes care in India during COVID-19 pandemic and national lockdown period: Guidelines for physicians. *Diabetes Metab. Syndr. Clin. Res. Rev.* **2020**, *14*, 273–276. [CrossRef]
19. Guo, S.; Guo, X.; Fang, Y.; Vogel, D. How doctors gain social and economic returns in online health-care communities: A professional capital perspective. *J. Manag. Inf. Syst.* **2017**, *34*, 487–519. [CrossRef]
20. Zhang, Y. *Affective-Discursive Practice in Online Medical Consultations in China: Emotional and Empathic Acts, Identity Positions, and Power Relations*; Springer Nature: Berlin/Heidelberg, Germany, 2022; Volume 11.
21. Khurana, S.; Qiu, L.; Kumar, S. When a doctor knows, it shows: An empirical analysis of doctors' responses in a Q&A forum of an online healthcare portal. *Inf. Syst. Res.* **2019**, *30*, 872–891.
22. Zhao, W.; Gu, D.; Yang, X.; Jia, M.; Liang, C.; Wang, X.; Zolotarev, O. MedT2T: An adaptive pointer constrain generating method for a new medical text-to-table task. *Future Gener. Comput. Syst.* **2024**, *161*, 586–600. [CrossRef]
23. Sabeeh, Z.; Syed Mustapha, S.; Mohamad, R. Healthcare knowledge sharing among a community of specialized physicians. *Cogn. Technol. Work.* **2018**, *20*, 105–124. [CrossRef]
24. Gu, D.; Wang, Q.; Chai, Y.; Yang, X.; Zhao, W.; Li, M.; Zolotarev, O.; Xu, Z.; Zhang, G. Identifying the Risk Factors of Allergic Rhinitis Based on Zhihu Comment Data Using a Topic-Enhanced Word-Embedding Model: Mixed Method Study and Cluster Analysis. *J. Med. Internet Res.* **2024**, *26*, e48324. [CrossRef]
25. Gu, D.; Liu, H.; Zhao, H.; Yang, X.; Li, M.; Liang, C. A deep learning and clustering-based topic consistency modeling framework for matching health information supply and demand. *J. Assoc. Inf. Sci. Technol.* **2024**, *75*, 152–166. [CrossRef]
26. Yan, Z.; Wang, T.; Chen, Y.; Zhang, H. Knowledge sharing in online health communities: A social exchange theory perspective. *Inf. Manag.* **2016**, *53*, 643–653. [CrossRef]
27. Yang, H.; Du, H.S.; He, W.; Qiao, H. Understanding the motivators affecting doctors' contributions in online healthcare communities: Professional status as a moderator. *Behav. Inf. Technol.* **2021**, *40*, 146–160. [CrossRef]

28. Charband, Y.; Jafari Navimipour, N. Online knowledge sharing mechanisms: A systematic review of the state of the art literature and recommendations for future research. *Inf. Syst. Front.* **2016**, *18*, 1131–1151. [CrossRef]
29. Yang, X.; Xi, N.; Gu, D.; Liang, C.; Liu, H.; Tang, H.; Hamari, J. Medical practice in gamified online communities: Longitudinal effects of gamification on doctor engagement. *Inf. Manag.* **2024**, *61*, 103906. [CrossRef]
30. Schroeder, D.A.; Graziano, W.G. The field of prosocial behavior: An introduction and overview. In *The Oxford Handbook of Prosocial Behavior*; Oxford University Press: Oxford, UK, 2015; pp. 3–34.
31. Dovidio, J.F.; Piliavin, J.A.; Schroeder, D.A.; Penner, L.A. *The Social Psychology of Prosocial Behavior*; Psychology Press: London, UK, 2017.
32. Aknin, L.B.; Whillans, A.V.; Norton, M.I.; Dunn, E.W. Happiness and prosocial behavior: An evaluation of the evidence. *World Happiness Rep.* **2019**, *2019*, 67–86.
33. Penner, L.A.; Dovidio, J.F.; Piliavin, J.A.; Schroeder, D.A. Prosocial behavior: Multilevel perspectives. *Annu. Rev. Psychol.* **2005**, *56*, 365–392. [CrossRef]
34. Barasch, A.; Levine, E.E.; Berman, J.Z.; Small, D.A. Selfish or selfless? On the signal value of emotion in altruistic behavior. *J. Personal. Soc. Psychol.* **2014**, *107*, 393. [CrossRef] [PubMed]
35. Majdandžić, J.; Amashaufer, S.; Hummer, A.; Windischberger, C.; Lamm, C. The selfless mind: How prefrontal involvement in mentalizing with similar and dissimilar others shapes empathy and prosocial behavior. *Cognition* **2016**, *157*, 24–38. [CrossRef]
36. Berman, J.Z.; Silver, I. Prosocial behavior and reputation: When does doing good lead to looking good? *Curr. Opin. Psychol.* **2022**, *43*, 102–107. [CrossRef]
37. Van Lange, P.A. Does empathy trigger only altruistic motivation? How about selflessness or justice? *Emotion* **2008**, *8*, 766. [CrossRef] [PubMed]
38. Swap, W.C. When prosocial behavior becomes altruistic: An attributional analysis. *Curr. Psychol.* **1991**, *10*, 49–64. [CrossRef]
39. Liu, Y.; Jiang, N.; Sun, H. The relationship between moral sense of life and prosocial behavior among medical students: The multiple mediating effects of meaning in life and empathy. *Chin. J. Behav. Med. Brain Sci.* **2021**, 944–948.
40. Meng, F.; Zhang, X.; Liu, L.; Ren, C. Converting readers to patients? From free to paid knowledge-sharing in online health communities. *Inf. Process. Manag.* **2021**, *58*, 102490. [CrossRef]
41. Meng, F.; Liu, Y.; Zhang, X.; Liu, L. General knowledge-sharing and patient engagement in online health communities: An inverted U-shaped relationship. *J. Knowl. Manag.* **2024**, *28*, 763–788. [CrossRef]
42. Ahluwalia, R.; Unnava, H.R.; Burnkrant, R.E. The moderating role of commitment on the spillover effect of marketing communications. *J. Mark. Res.* **2001**, *38*, 458–470. [CrossRef]
43. Song, W.; Chen, J.; Li, W. Spillover effect of consumer awareness on third parties' selling strategies and retailers' platform openness. *Inf. Syst. Res.* **2020**, *32*, 172–193. [CrossRef]
44. Bettels, J.; Wiedmann, K.-P. Brand logo symmetry and product design: The spillover effects on consumer inferences. *J. Bus. Res.* **2019**, *97*, 1–9. [CrossRef]
45. Erreygers, S.; Vandebosch, H.; Vranjes, I.; Baillien, E.; De Witte, H. Feel good, do good online? Spillover and crossover effects of happiness on adolescents' online prosocial behavior. *J. Happiness Stud.* **2019**, *20*, 1241–1258. [CrossRef]
46. Olson, E.L. The impact of intra-brand platform sharing on brand attractiveness. *J. Prod. Brand Manag.* **2009**, *18*, 212–217. [CrossRef]
47. Smith, D.C.; Park, C.W. The effects of brand extensions on market share and advertising efficiency. *J. Mark. Res.* **1992**, *29*, 296–313. [CrossRef]
48. Balachander, S.; Ghose, S. Reciprocal spillover effects: A strategic benefit of brand extensions. *J. Mark.* **2003**, *67*, 4–13. [CrossRef]
49. Pontes, N.; Pontes, V. Spillover effects of competitive rivalry on brand extensions. *J. Brand Manag.* **2021**, *28*, 402–412. [CrossRef]
50. Tezer, A.; Tofighi, M. CSR spillover effect: The influence of a brand's corporate social responsibility activity on competing brands. *J. Mark. Manag.* **2021**, *37*, 651–670. [CrossRef]
51. Chae, I.; Stephen, A.T.; Bart, Y.; Yao, D. Spillover effects in seeded word-of-mouth marketing campaigns. *Mark. Sci.* **2017**, *36*, 89–104. [CrossRef]
52. Zou, P.; Li, G. How emerging market investors' value competitors' customer equity: Brand crisis spillover in China. *J. Bus. Res.* **2016**, *69*, 3765–3771. [CrossRef]
53. Seo, S.; Jang, S.S. Who spoils the barrel? Negative spillover effect on competitor brands during food crises. *Int. J. Hosp. Manag.* **2021**, *98*, 103047. [CrossRef]
54. Islam, S.; Grigoriadis, C.; Papastergiou, S. Information sharing for creating awareness for securing healthcare ecosystem. In Proceedings of the 2023 19th International Conference on the Design of Reliable Communication Networks (DRCN), Vilanova i la Geltrú, Spain, 22–25 March 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5.
55. Xu, J.; Jain, H.K.; Gu, D.; Liang, C. Business-Process-Driven Service Composition in a Hybrid Cloud Environment. *Inf. Syst. Front.* **2025**, *27*, 259–281. [CrossRef]

56. Pikkarainen, M.; Kemppainen, L.; Xu, Y.; Jansson, M.; Ahokangas, P.; Koivumäki, T.; Hong Gu, H.; Francis Gomes, J. Resource integration capabilities to enable platform complementarity in healthcare service ecosystem co-creation. *Balt. J. Manag.* **2022**, *17*, 688–704. [CrossRef]
57. Zhou, P.; Xie, Y.; Liang, C.; Zhu, J.; Zhao, S. The effect of prosocial behavior and its intensity on doctors' performance in an online health community. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 874. [CrossRef]
58. Francetic, I.; Meacock, R.; Elliott, J.; Kristensen, S.R.; Britteon, P.; Lugo-Palacios, D.G.; Wilson, P.; Sutton, M. Framework for identification and measurement of spillover effects in policy implementation: Intended non-intended targeted non-targeted spillovers (INTENTS). *Implement. Sci. Commun.* **2022**, *3*, 30. [CrossRef] [PubMed]
59. Gong, W.; Tang, J.; Chen, Y.; Ma, J. Collaborative Communication: A Qualitative Study of Roles and Emphases of Health Care Providers in Obstetrics and Gynecology. *J. Multidiscip. Healthc.* **2024**, *17*, 1913–1922. [CrossRef]
60. Deng, Z.; Deng, Z.; Liu, S.; Evans, R. Knowledge transfer between physicians from different geographical regions in China's online health communities. *Inf. Technol. Manag.* **2023**, *26*, 83–100. [CrossRef]
61. Wu, M.; Brandhorst, H.; Marinescu, M.C.; Lopez, J.M.; Hlava, M.; Busch, J. Automated metadata annotation: What is and is not possible with machine learning. *Data Intell.* **2023**, *5*, 122–138. [CrossRef]
62. Freeman, M.; Savva, N.; Scholtes, S. Economies of scale and scope in hospitals: An empirical study of volume spillovers. *Manag. Sci.* **2021**, *67*, 673–697. [CrossRef]
63. Basajja, M.; Nambobi, M.; Wolstencroft, K. Possibility of enhancing digital health interoperability in Uganda through FAIR data. *Data Intell.* **2022**, *4*, 899–916. [CrossRef]
64. Xu, Z.; Liu, X.; Meng, L.; Lyu, X. More knowledge, more choices? How peer recognition of physicians' knowledge sharing affect patients' consultation in online health communities. *Front. Public Health* **2024**, *12*, 1376887. [CrossRef]
65. Zhang, X.; Dong, X.; Xu, X.; Guo, J.; Guo, F. What influences physicians' online knowledge sharing? A stimulus–response perspective. *Front. Psychol.* **2022**, *12*, 808432. [CrossRef] [PubMed]
66. Lu, J.; Bai, J.; Zhao, H.; Zhang, X. The effect of "offline-to-online" trust transfer on the utilization of online medical consultation among Chinese rural residents: Experimental study. *J. Med. Internet Res.* **2023**, *25*, e43430. [CrossRef] [PubMed]
67. Morgan, C.J. Reducing bias using propensity score matching. *J. Nucl. Cardiol.* **2018**, *25*, 404–406. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

systems

MDPI

Article

Lagged Stance Interactions and Counter-Spiral of Silence: A Data-Driven Analysis and Agent-Based Modeling of Technical Public Opinion Events

Kaihang Zhang [1,2,†], Changqi Dong [1,2,*], Yifeng Guo [2,*], Wuai Zhou [2,†], Guang Yu [1] and Jianing Mi [2,3]

[1] School of Management, Harbin Institute of Technology, Harbin 150001, China; 24b310006@stu.hit.edu.cn (K.Z.); yug@hit.edu.cn (G.Y.)
[2] Harbin Institute of Technology-China Mobile Limited 5G Application Innovation Joint Research Institute, Harbin 150006, China; zhouwuai@cmict.chinamobile.com (W.Z.); mijianing@hit.edu.cn (J.M.)
[3] School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China
* Correspondence: 21b910041@stu.hit.edu.cn (C.D.); guoyifeng@cmict.chinamobile.com (Y.G.)
† These authors contributed equally to this work.

Abstract: Understanding the dynamics of public opinion formation in digital environments is crucial for managing technological communications effectively. This study investigates stance interactions and opinion reversal phenomena in technical discourse through analysis of the Manus AI controversy that generated approximately 36,932 social media interactions during March 2025. Employing an integrated methodology combining Large Language Model (LLM)-enhanced stance detection with agent-based modeling (ABM), we reveal distinctive patterns challenging traditional public opinion theories. Our cross-correlation analysis identifies significant lagged interaction effects between skeptical and supportive stances, demonstrating how critical expressions trigger amplified counter-responses rather than inducing silence. Unlike prior conceptualizations of counter-silencing that emphasize ideological resistance or echo chambers, our notion of the "counter-spiral of silence" specifically highlights lagged emotional responses and reactive amplification triggered by minority expressions in digital technical discourse. We delineate its boundary conditions as arising under high emotional salience, asymmetrical expertise, and platform structures that enable real-time feedback. The agent-based simulation reproduces empirical patterns, revealing how emotional contagion and network clustering mechanisms generate "counter-spiral of silence" phenomena where challenges to dominant positions ultimately strengthen rather than weaken those positions. These findings illuminate how cognitive asymmetries between public expectations and industry realities create distinctive discourse patterns in technical contexts, offering insights for managing technology communication and predicting public response trajectories in rapidly evolving digital environments.

Keywords: spiral of silence; agent-based modeling; stance detection; emotional contagion; large language models; technological nationalism; public opinion dynamics

1. Introduction

In recent years, the digital landscape has witnessed an unprecedented proliferation of technical discourse events that rapidly evolve through complex patterns of opinion formation, stance reversal, and emotional contagion [1]. These phenomena represent significant challenges for traditional public opinion theories, as they often exhibit dynamics that

transcend conventional understanding of how public discourse unfolds [2–4]. The Manus technology controversy of March 2025 presents a compelling case study of such complex dynamics, where an AI product initially celebrated as a breakthrough innovation rapidly faced technical scrutiny and public backlash, creating a dramatic stance reversal within a remarkably short timeframe.

The classical Spiral of Silence theory, first proposed by Noelle-Neumann [5], suggests that individuals who perceive their opinions to be in the minority tend to remain silent due to fear of isolation. However, contemporary digital environments appear to foster what scholars have begun to identify as a "Counter-Spiral of Silence" phenomenon, wherein perceived minority opinions can gain surprising momentum through lagged emotional interactions and digital amplification mechanisms [6,7]. These dynamics are particularly evident in technical discourse events, where specialized knowledge, information asymmetries, and rapid digital diffusion create unique conditions for public opinion formation [8].

The intersection of technological discourse with nationalist sentiment—often characterized as "technological nationalism"—further complicates these dynamics in distinctive ways. Technological nationalism, wherein technological achievements become proxies for national pride and identity, creates a charged discursive environment where stance expressions carry additional emotional and identity-based dimensions beyond mere technical assessments. In such contexts, lagged interactions between skeptical and supportive stances are particularly pronounced, as challenges to technological claims may be perceived not simply as technical criticisms but as challenges to collective identity narratives. This creates asymmetric response patterns where skeptical expressions trigger disproportionate counter-mobilization from supporters who are motivated by both technical and identity-based concerns. The Manus controversy exemplifies this dynamic, where initial technical scrutiny quickly evolved into broader discourse about national innovation capabilities and competitive positioning in global AI development, creating distinctive temporal patterns in stance expression and emotional intensification that traditional spiral of silence theory struggles to explain.

The emergence of new computational methodologies—particularly Large Language Models (LLMs) and Agent-Based Modeling (ABM)—offers unprecedented opportunities to analyze these complex phenomena with greater precision [9,10]. While traditional sentiment analysis based on BERT models has provided valuable insights into public opinion dynamics [11,12], LLMs demonstrate superior capabilities in capturing nuanced stances and emotional states across diverse discourse contexts [13]. Similarly, ABM approaches allow researchers to simulate the microinteractions that generate emergent patterns of opinion formation and diffusion [14,15], providing a computational framework for testing theoretical propositions about digital discourse dynamics.

This paper presents a comprehensive data-driven analysis of stance interactions and counter-silence phenomena in technical public opinion events, using the Manus controversy as a focal case study. By integrating advanced stance detection through LLM-based classification with agent-based simulation of opinion dynamics, we aim to address three central research questions: (1) How do lagged stance interactions manifest in technical public opinion events? (2) What mechanisms drive the counter-spiral of silence phenomenon in digital environments? (3) How can agent-based modeling illuminate the microprocesses that generate macro-level patterns of opinion reversal?

Our investigation follows a three-phase methodological approach. First, we conduct a descriptive statistical analysis of stance distribution and temporal evolution across the Manus controversy timeline, identifying key inflection points in public discourse. Second, we employ both traditional BERT models and advanced LLMs to classify stance positions and emotional expressions in social media content, demonstrating the superior perfor-

mance of LLMs in capturing stance nuances. Third, we develop an agent-based model that incorporates emotional dynamics, network effects, and silence thresholds to simulate the emergence of opinion patterns observed in the empirical data.

The significance of this research extends beyond academic interest in digital discourse dynamics. Understanding the mechanisms that drive stance reversal and counter-silence phenomena has important implications for technology communication, crisis management, and digital platform governance [16,17]. By illuminating how lagged emotional interactions influence public discourse trajectories, this research contributes to a more sophisticated theoretical framework for analyzing and potentially anticipating the evolution of public opinion in contemporary digital environments [18].

The remainder of this paper is structured as follows. Section 2 reviews relevant literature on the Spiral of Silence theory, stance detection in technical discourse, and agent-based approaches to social media dynamics. Section 3 details our data collection, model comparison methodology, and agent-based model design. Section 4 presents our empirical findings regarding temporal patterns, lagged stance interactions, and simulation results. Section 5 discusses theoretical and practical implications, while Section 6 concludes with contributions, limitations, and directions for future research.

2. Literature Review

2.1. Theoretical Foundations of Spiral of Silence and Counter-Silence Phenomena

The Spiral of Silence theory, first articulated by Elisabeth Noelle-Neumann in 1974 [5], represents one of the most influential frameworks for understanding the dynamics of public opinion formation. This theory posits that individuals continuously monitor their social environment to assess the distribution of opinions and, when perceiving themselves to hold minority views, often choose silence over expression due to fear of social isolation [5,19]. Noelle-Neumann argues that this self-censorship creates a spiraling process wherein perceived majority opinions gain disproportionate visibility while minority views become increasingly suppressed, eventually leading to a distorted perception of public opinion distribution—a phenomenon termed "pluralistic ignorance" [20].

The theory's central mechanism relies on what Noelle-Neumann termed the "quasi-statistical sense"—an intuitive ability to gauge opinion climates through environmental observation [21]. Traditional mass media, according to this framework, plays a critical role in amplifying this spiral effect by presenting opinions disproportionately to their actual prevalence in society [22]. This creates a self-reinforcing cycle where perceived majority opinions gain further prominence while alternative viewpoints become increasingly marginalized in public discourse [23].

However, the digital transformation of communication environments has complicated these dynamics significantly. Several scholars have identified phenomena that appear to contradict or modify the classical spiral of silence predictions in online contexts [8,24]. Neubaum and Krämer [6] found that social media users often demonstrated reduced willingness to discuss controversial topics both online and offline, suggesting a potential amplification of silence dynamics across communicative domains. Conversely, Brady et al. [25] documented instances where social media platforms appeared to facilitate the expression of minority opinions, particularly when users perceived themselves within supportive "echo chambers" or when platform affordances enabled anonymous expression.

These divergent findings have led to the conceptualization of what some scholars term the "Counter-Spiral of Silence" phenomenon [6,7]. This concept suggests that digital environments can, under certain conditions, reverse the traditional silencing mechanism by enabling minority opinions to gain rapid visibility and momentum through various tech-

nological affordances and network effects [26]. Several mechanisms have been proposed to explain this counter-spiral effect, including reduced fear of isolation in anonymous online settings [27], algorithmic amplification of emotionally charged content [25], and the formation of ideologically homogeneous communities that provide psychological safety for expressing controversial views [28].

The temporal dimension of these dynamics has received increasing scholarly attention, with recent research suggesting that opinion expression in digital environments often follows complex non-linear patterns characterized by lagged interactions between different stance positions [29]. Fox and Holt [7] propose that perceived affordances of communication channels significantly influence whether individuals express their opinions or self-censor, suggesting that technological features of platforms interact with psychological processes to shape expression behaviors. These temporal and technological dimensions provide crucial context for understanding how stance dynamics unfold in contemporary digital discourse.

A more systematic unpacking of the Spiral of Silence theory reveals three interdependent components at its core: (1) the quasi-statistical sense, or individuals' perceived awareness of dominant social opinion climates; (2) the fear of isolation, whereby individuals refrain from expressing minority views to avoid social sanction; and (3) the dynamic spiraling process, where the perceived dominance of one opinion reduces the visibility of opposing views, thereby further reinforcing silence and public misperception. These elements function as a self-reinforcing loop, particularly potent in centralized mass media systems. However, digital environments—with their affordances for anonymity, rapid feedback, and algorithmic amplification—challenge each of these assumptions by disrupting visibility hierarchies, lowering isolation risks, and enabling reactive expression. Recognizing these divergences is essential for understanding the emergence of counter-silence dynamics observed in this study.

2.2. Sentiment Analysis and Stance Detection in Technical Public Opinion

Computational approaches to analyzing public opinion have evolved substantially, transitioning from simple lexicon-based sentiment analysis to sophisticated stance detection methods capable of discerning nuanced positions on complex issues [30,31]. This evolution is particularly relevant for technical discourse analysis, where specialized terminology, domain-specific knowledge, and multifaceted argumentative structures require advanced computational techniques [32].

Traditional sentiment analysis methods typically classify text along a positive-negative continuum, often supplemented with a neutral category [33]. While valuable for broad assessments of emotional valence, these approaches frequently struggle to capture the nuanced stance positions that characterize technical discourse [34]. Stance detection, in contrast, aims to identify the author's position toward a particular target—whether they support, oppose, or remain neutral toward a specific entity, concept, or claim [31]. This distinction is particularly important in technical controversies, where expressions of skepticism, qualified support, or domain-specific critiques may not align neatly with simple sentiment categories [35].

The emergence of transformer-based models, particularly BERT (Bidirectional Encoder Representations from Transformers) and its variants, has significantly advanced the field of stance detection [11,12]. These models leverage deep contextual representations and massive pre-training datasets to capture semantic nuances in text, enabling more accurate classification of complex stance positions [36]. In technical discourse analysis, specialized BERT variants such as SciBERT [37] and RoBERTa-wwm [38] have

demonstrated improved performance by incorporating domain-specific training data and structural knowledge.

Most recently, Large Language Models (LLMs) have pushed computational stance analysis boundaries even further [13]. Models such as GPT-4, Claude, and GLM-4 leverage their scale and sophisticated architecture to capture increasingly subtle stance positions, outperforming traditional BERT-based approaches on various benchmark tasks [39]. Recent comparative studies demonstrate that LLMs exhibit superior performance in identifying complex stance categories, particularly in highly nuanced technical discussions [40]. This advantage stems from their ability to process longer contextual windows, incorporate world knowledge, and recognize domain-specific reasoning patterns [41].

Despite these advances, several challenges persist in computational stance analysis of technical discourse. These include the domain-transfer problem, where models trained on general language corpora struggle with specialized technical terminology [42]; the context-dependency challenge, where stance interpretation requires understanding broader disciplinary or societal debates [43]; and the cultural-linguistic variation issue, where stance expressions differ substantially across languages and cultural contexts [44]. These challenges highlight the need for continued methodological innovation at the intersection of computational linguistics and domain-specific knowledge.

2.3. Agent-Based Modeling Approaches in Social Media Opinion Dynamics

Agent-based modeling (ABM) has emerged as a powerful computational approach for investigating complex social phenomena that emerge from individual-level interactions [14,45]. In the context of social media opinion dynamics, ABM enables researchers to explore how micro-level behaviors, such as information sharing, opinion expression, and social influence, generate macro-level patterns of opinion distribution and evolution [15,46].

The foundational premise of ABM in this domain is that social media users can be represented as computational agents with specified attributes (e.g., initial opinions, expression thresholds, network positions) and behavioral rules (e.g., interaction tendencies, updating mechanisms) [47]. By simulating interactions among these agents over time, researchers can observe emergent patterns that might explain empirically observed phenomena in digital discourse [48]. This "generative social science" approach, as characterized by Epstein [49], allows for rigorous testing of theoretical mechanisms that would be difficult or impossible to isolate in observational studies.

ABM applications to social media opinion dynamics have explored various mechanisms related to the Spiral of Silence theory and its digital modifications. Sohn and Geidner [50] developed models incorporating psychological thresholds for opinion expression, demonstrating how small variations in these thresholds can lead to dramatically different macro-level opinion distributions. Similarly, Xiong and Liu [51] explored how network homophily interacts with expression dynamics, showing that highly clustered networks tend to produce fragmented opinion landscapes with reduced central silencing effects.

More recent ABM approaches have incorporated increasingly sophisticated psychological and social mechanisms. Stefanelli and Seidl [52] integrated emotional contagion dynamics into opinion expression models, demonstrating how affective responses can accelerate or inhibit the formation of silence spirals. Similarly, Acerbi [53] incorporated concepts from cultural evolution theory to model how stance positions evolve through complex social learning mechanisms in digital networks.

A critical methodological advancement in this domain is the development of empirically calibrated ABMs, where model parameters are estimated from observed data rather than set arbitrarily [54]. This approach increases the ecological validity of simulations

and enables more direct comparison between model predictions and empirical observations [55]. For instance, Bail et al. [56] used digital trace data to calibrate agent attributes and network structures in their models of political polarization dynamics, demonstrating how empirically grounded ABMs can generate insights about causal mechanisms underlying observed phenomena.

The integration of ABM with machine learning techniques represents another frontier in this domain [57]. Recent work has used machine learning to identify patterns in empirical data that inform agent rule specifications [58], optimize model parameters for improved fit with observed phenomena [59], and create hybrid modeling approaches that combine data-driven and theory-driven elements [60]. These methodological innovations enable increasingly sophisticated simulations of how stance dynamics evolve in complex digital environments.

2.4. Recent Advances Supporting the Counter-Spiral of Silence Framework

Recent scholarship across computational social science, public opinion theory, and affective computing has begun to reconceptualize the Spiral of Silence (SoS) framework in response to the affordances and dynamics of digital platforms. In contrast to the classical assumption that minority voices tend to fall silent, emerging studies indicate that under specific emotional, symbolic, and algorithmic conditions, such voices may function as catalysts of majority response, forming a counter-spiral dynamic.

From a computational perspective, Wei et al. [39] demonstrate that Chain-of-Thought prompting significantly enhances large language models' (LLMs) reasoning capacity, including in stance-sensitive contexts, while Yang et al. [40] propose inductive reasoning frameworks to model rule emergence from factual discourse. These advances not only improve stance detection performance but also enable LLMs to trace the implicit logic of emotional and rhetorical progression in minority expressions.

In a broader socio-cognitive context, Burton et al. [45] argue that LLMs fundamentally reshape collective intelligence by altering the way discourse is aggregated and amplified, calling for new frameworks to understand emergent opinion spirals. Similarly, Zhang et al. [61] provide a comprehensive survey on stance detection, emphasizing how LLM-driven sentiment and opinion tracking can expose hidden dynamics of polarization and alignment in social media discourse.

These computational insights align with findings in social communication. Sohn [62] uses agent-based modeling to show that the traditional spiral of silence is unlikely to dominate unless media narratives are highly homogeneous and networks hyperconnected, indicating that fragmented visibility may hinder rather than sustain silence. In support of this, Laor [63] finds that closed-topic Facebook groups offer emotional refuge and expressive safety for dissenters, often leading them to reintegrate their opinions into broader discourse, thereby reversing the spiral. Furthermore, Kuncoro and Hasanah [64] analyze how algorithmic reinforcement within echo chambers can amplify emotionally resonant radical views, highlighting that visibility and engagement metrics, not majority consensus, often drive attention hierarchies and emotional contagion.

Finally, Chen [65] models information diffusion under diverse topologies, showing how structurally peripheral nodes can, under the right emotional payload and timing, initiate cascades that rival those originating from central nodes, supporting the notion that emotional intensity, rather than structural centrality alone, determines influence.

Taken together, these studies offer empirical and theoretical support for our proposition of a Counter-Spiral of Silence: a dynamic in which emotionally charged minority expressions trigger delayed but amplified majority reactions, particularly in techno-political controversies. This literature suggests that under high emotional arousal, symbolic sensi-

tivity, and algorithmically mediated visibility, minority positions act not as suppressible noise but as ignition points of expressive counter-mobilization.

3. Data and Methods

3.1. Data Collection

This research focuses on the social media public opinion triggered by the Manus technology controversy, systematically collecting high-quality opinion text data. The data collection period was determined to be from 6 March to 10 March 2025, spanning five days. This time window, as indicated by Baidu Index data, captured a complete cycle from "Manus product announcement" to the controversy's peak and subsequent stabilization, allowing for a comprehensive analysis of the technical public opinion from emergence to peak and cooling phases.

Data collection primarily focused on the Weibo platform, which features a diverse user base and active interactions, providing a relatively comprehensive reflection of public perception and evaluation of Manus. The collection process utilized custom-developed Python 3.8-based web crawlers combined with Weibo's open API, using "Manus", "Monica.im", and "AI Agent" as search keywords to obtain microblog texts containing these terms. To ensure data completeness and representativeness, multiple strategies were employed: on the one hand, an incremental collection method was implemented, with data crawling every 4 h to accurately capture real-time changes in public opinion; on the other hand, in-depth crawling was conducted for popular microblogs (those with repost counts, comment counts, or like counts exceeding 1000) to obtain their complete textual content and interaction information.

The data cleaning and preprocessing stage primarily included: first, deduplication of the original data, removing completely identical reposted content; second, filtering advertisements, batch content published by bot accounts, and comments consisting solely of emoji without substantial content; third, standardizing text formats, processing special characters, URLs, and non-standard expressions. After these processes, a total of 36,932 valid Weibo text entries were obtained, constituting the core dataset for this research. To better illustrate the qualitative texture of these data, we present in Table 1 a selection of representative posts categorized by stance, along with corresponding wordclouds that visualize dominant expressions within each group.

To ensure the reliability of subsequent analyses, user attribute data (including follower count, following count, user verification type, etc.) and interaction data (including repost count, comment count, like count, etc.), as well as timestamp information, were also collected to construct temporal evolution models and social network structures. These auxiliary data help characterize the heterogeneous features and interaction patterns of opinion subjects, providing an empirical foundation for agent-based modeling. A summary of key user attributes and stance distributions across different account types is provided in Table 2, offering a descriptive overview of the heterogeneous participant landscape in this controversy.

Table 1. Representative public reactions to the Manus controversy with stance-based wordclouds.

Stance	Representative Sample Text	Wordcloud
Supportive	NB, 没想到这么快就来了 manus或许会将AI推向新高度 #人工智能[超话]##人工智能##manus# (Great, I didn't expect it to arrive so soon. Manus may push AI to new heights. # Artificial Intelligence [Super Talk] # # Artificial Intelligence # # manus#)	
	Manus, 长上下文通用ai agent, [doge]这完全不亚于Deepseek, 太强了, 感觉能直接改变人机交互方式。 (Manus, Long context universal AI agent, [doge] this is no less powerful than Deepseek, it feels like it can directly change the way human-machine interaction is done.)	
	Manus激发了人工智能领域的新活力，直接操作电脑完成复杂工作, 令人期待 (Manus has stimulated new vitality in the field of artificial intelligence, enabling direct operation of computers to complete complex tasks, which is highly anticipated)	
	这个ai智能体manus突然爆火！比如跟他说，帮我生成一个小米su7的介绍的10页ppt, 然后它就会自己主动去搜索su7的介绍，并准备ppt的大纲，设计ppt的模板, 并自动撰写ppt！ 全程不用再输入。。 (This AI intelligent agent manus suddenly went viral! For example, tell him to help me generate a 10 page PPT introducing the Xiaomi Su7, and then it will actively search for the introduction of Su7, prepare the outline of the PPT, design the PPT template, and automatically write the PPT! No need to input again throughout the process..)	
	//@股市小杨哥:又一家中国科技公司一夜爆红，Manus, 搞出了中国首个通用AI智能体，应该是目前全世界最牛逼的通用AI智能体。 (//@Stock market Xiao Yang Ge: Another Chinese technology company went viral overnight, Manus, China has developed its first universal AI intelligent agent, which should be the most powerful universal AI intelligent agent in the world at present.)	

Table 1. *Cont.*

Stance	Representative Sample Text	Wordcloud
Skeptical/Oppositional	我算是看出来了，这manus的文章都是商单。btw, 他本身不是个大模型，是基于大模型的应用，怎么配跟openai、deepseek比呢？[笑cry] (I can see that all the articles in this manual are commercial orders. btw, He is not a big model himself, but an application based on big models. How can he compare with OpenAI and Deepseek? [Laughing cry])	
	Manus花了多少钱找人营销？我仔细看了看demo，并没有达到agent的标准，更像是个工具箱，这些AI博主真是拿钱就乱吹，还扯上“国运级AI”。。。。。 (How much did Manus spend on marketing? I carefully looked at the demo and it didn't meet the standards of an agent. It was more like a toolbox, and these AI bloggers were just bragging about it with money, even bringing up “national level AI”.....)	
	#Manus为何一夜爆火#只看到软广式的炒作[汗]没看到所谓的火！[666] (#Why did Manus go viral overnight? I only saw soft advertising hype [sweat] and didn't see the so-called fire! [666])	
	“尽管 Manus尚未全面开放, 但其热度已席卷全网”。至少宣传是铺天盖地, 但是骡子是马还得出来遛一遛 (Although Manus has not yet been fully opened, its popularity has swept across the entire internet. At least the promotion is overwhelming, but mules are horses and still need to come out for a walk)	
	Manus这么牛逼吗？为何能一夜爆火? (Is Manus so awesome? Why did it explode overnight?)	

Table 2. Descriptive characteristics of users expressing skeptical and supportive stances toward Manus.

User Type	Skeptical/Denial	Supportive
Followers (avg.)	42,862	19,523
Followings (avg.)	348	514
Top-tier Influencers	0.0173	0.0094
Corporate Accounts	0.0239	0.0209
Entry-level Influencers	0.0015	0.0009
Verified Celebrities	0.3009	0.1984
Organizations	0.0673	0.0392
Media	0.1480	0.1271
Government Accounts	0.0054	0.0061
Ordinary Users	0.4249	0.5407
Unknown Type	0.0156	0.0226
Campus Accounts	0.0000	0.0005

3.2. Comparison of LLM and BERT Models in Stance and Sentiment Detection

To effectively analyze the complex stance positions and emotional expressions in the Manus controversy dataset, we conducted a comprehensive evaluation of different language models' performance in both stance recognition and sentiment analysis tasks. Our comparison included two Large Language Models (LLMs)—GLM-4-Flash and DeepSeek-V3—and three variants of BERT models—BERT-base-chinese, RoBERTa-wwm, and MacBERT. This comparative analysis aimed to identify the most effective approach for processing large volumes of technical public opinion text data.

For the stance recognition task, we employed a three-category classification framework (stance-3) to identify whether a post expressed support for Manus (Bullish), skepticism about its capabilities (Skeptical), or outright rejection of its claims (Deny). For sentiment analysis, we implemented a more granular eight-category classification system (sentiment-8) encompassing primary emotions (joy, anger, sadness, fear) and secondary emotional states (anticipation, surprise, trust, disgust) based on established emotion classification frameworks.

To ensure rigorous evaluation, we manually annotated a subset of 2000 Weibo posts selected through stratified random sampling to maintain temporal and thematic representation. Three independent annotators with background knowledge in both AI technology and communication studies performed the annotation, achieving an inter-annotator agreement of 0.81 (Cohen's Kappa). This annotated dataset was then split into training (70%), validation (15%), and testing (15%) sets for model development and evaluation.

As shown in Table 3, our evaluation revealed substantial performance differences between LLM-based classifiers and traditional BERT models. The LLM approaches demonstrated notably superior performance, with DeepSeek-V3 achieving the highest accuracy in both stance recognition (0.83) and sentiment analysis (0.80) tasks. Among the BERT variants, RoBERTa-wwm consistently outperformed both the basic BERT-base-Chinese model and MacBERT, achieving accuracies of 0.73 for stance recognition and 0.71 for sentiment analysis.

Table 3. Performance of different classifiers, such as LLM and the classic BERT model, in stance recognition and sentiment analysis.

Model Type	Model Name	Features	Stance-3		Sentiment-8	
			Accuracy	F1-Score	Accuracy	F1-Score
LLM	GLM-4-Flash	Free, long-context	0.81	0.74	0.79	0.66
	DeepSeek-V3	Mixture-of-Experts	0.83	0.75	0.80	0.71
Bert	BERT-base-Chinese	Basic Chinese BERT	0.65	0.51	0.62	0.53
	RoBERTa-wwm	RoBERTa with Whole Word Masking	0.73	0.67	0.71	0.62
	MacBERT	Adversarial training optimization	0.72	0.68	0.75	0.67

The performance gap between LLMs and BERT models was particularly pronounced in handling complex stance positions and nuanced emotional expressions. For instance, in the stance recognition task, DeepSeek-V3 achieved an F1-score of 0.75 compared to BERT-base-Chinese's 0.51, representing a 47% improvement. This substantial performance differential can be attributed to LLMs' enhanced capability to interpret contextual cues, technical terminology, and implicit stance signals present in the data.

The superior performance of LLMs over BERT models can be attributed to several specific advantages in processing technical discourse. First, LLMs benefit from substantially larger context windows (32K tokens for DeepSeek-V3 compared to 512 tokens for BERT), enabling them to process entire posts with surrounding conversation threads rather than isolated text fragments. This is particularly important for technical controversies where stance positions often depend on broader contextual references. Second, LLMs demonstrate superior capabilities in handling technical terminology and jargon through their massive pre-training exposure to specialized domains, allowing them to recognize stance signals in technically nuanced expressions where BERT models often misclassify technical criticism as neutral content.

The performance difference is most pronounced when processing three specific linguistic patterns common in technical discourse: (1) implicit stance expressions using domain-specific metaphors or analogies, where LLMs significantly outperform BERT models; (2) stance positions expressed through technical comparisons or benchmarking language, where traditional models struggle with the specialized vocabulary; and (3) emotionally nuanced technical assessments combining factual claims with subjective evaluations, which require the contextual understanding LLMs provide. This pattern suggests that LLMs' advantage increases with the technical complexity and implicit nature of the stance expressions, making them particularly valuable for analyzing technical controversies where opinions are often embedded within sophisticated technical arguments.

Figure 1 illustrates the training progression of the three BERT variants across ten epochs for both stance and sentiment classification tasks. The accuracy curves demonstrate that RoBERTa-wwm consistently outperformed the other BERT variants throughout the training process, with more rapid initial convergence and better generalization to testing data. However, all BERT models exhibited a notable gap between training and testing accuracies, suggesting limited generalization capabilities when applied to unseen data.

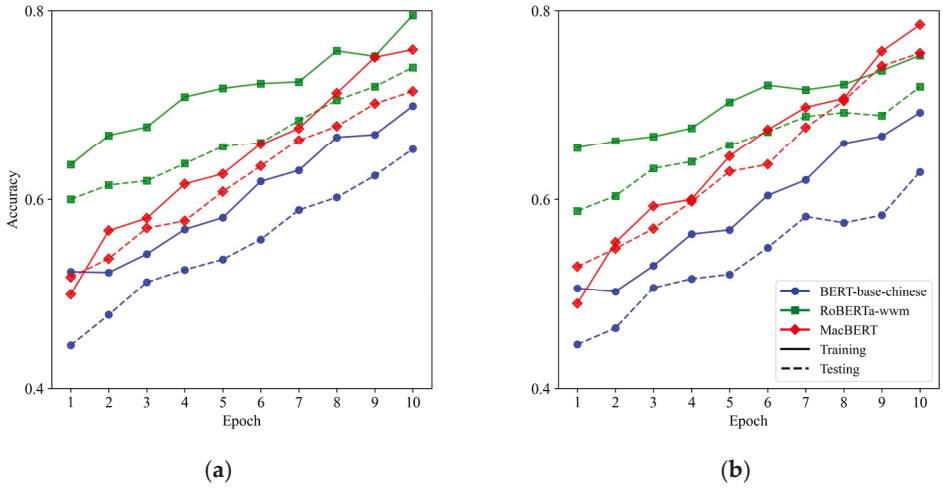


Figure 1. The accuracy change curve of the three pre-trained language models during the training epoch. (**a**) Stance. (**b**) Sentiment.

The improved performance of LLMs over BERT models can be attributed to several factors. First, LLMs benefit from significantly larger pre-training corpora and more complex model architectures, enabling them to capture deeper semantic relationships. Second, their mixture-of-experts architecture (in DeepSeek-V3) and long-context windows (in GLM-4-Flash) provide advantages in processing longer text segments with complex contextual dependencies. Third, LLMs demonstrate superior capability in adapting to domain-specific language patterns through in-context learning, which is particularly valuable when analyzing technical discourse.

Based on these comparative results, we selected DeepSeek-V3 as our primary classifier for processing the complete Manus controversy dataset. This model was applied to classify all 36,932 posts in our corpus, generating stance and sentiment labels that served as the foundation for subsequent temporal analysis of opinion evolution and lagged interaction effects. The resulting classified dataset enabled us to track how different stance positions and emotional expressions evolved throughout the controversy timeline, providing essential data for understanding the Counter-Spiral of Silence phenomenon observed in this technical public opinion event.

3.3. Agent-Based Model Design

To investigate the micro-level mechanisms driving lagged stance interactions and counter-spiral phenomena in technical discourse, we developed a comprehensive agent-based model simulating opinion formation dynamics within networked communication environments. The model integrates theoretical insights from complex systems theory, emotional contagion dynamics, and spiral of silence mechanisms to create a computational framework capable of reproducing the emergent patterns observed in empirical data.

Our agent-based simulation represents each social media participant as an autonomous agent characterized by both stance attributes and multidimensional emotional states. Each agent possesses an initial opinion value (ranging from -1 to 1, with positive values supporting Manus and negative values questioning it), individualized sensitivity thresholds (normally distributed around 0.7), and a seven-dimensional emotional state vector representing the intensity of discrete emotions (joy, sadness, anger, fear, surprise, disgust, and neutral). This design allows for modeling the complex interplay between cognitive positions and affective responses that characterizes real-world discourse.

To capture the heterogeneous structure of online communication networks, we implemented a hybrid network topology combining three complementary network generation

algorithms. The model incorporates a connected caveman graph (50 clusters of 10 nodes, totaling 500 nodes) representing tight-knit communities, a Barabási-Albert scale-free network (500 nodes with 3 edges per new node) capturing preferential attachment dynamics, and a Watts-Strogatz small-world network (500 nodes, 4 initial neighbors per node, rewiring probability of 0.1) modeling the characteristic combination of high clustering and short path lengths observed in social media environments. These networks were interconnected through strategic bridge connections (30 edges randomly created between subgraphs) to facilitate information flow across community boundaries, creating a realistic approximation of social media interaction structures. The total network size of 1500 nodes ensures sufficient topological complexity to fully manifest the power-law degree distribution characteristic of scale-free networks and the high clustering coefficient with short average path length features of small-world networks.

The behavioral mechanisms governing agent interactions implement core theoretical principles of the spiral of silence theory while extending them to account for digital communication affordances. During each simulation step, agents observe their local network neighborhood and assess the proportion of neighbors holding similar opinions. When this proportion exceeds their individualized sensitivity threshold (with small random variations to capture decision stochasticity), agents express their opinions and influence neighboring agents; otherwise, they remain silent and increment their silence duration counter. This threshold-based expression mechanism operationalizes the fundamental fear-of-isolation principle central to the Spiral of Silence theory.

A key innovation in our model is the introduction of dynamic opinion adjustment mechanisms that capture both silence-induced conformity and counter-spiral phenomena. Agents maintaining prolonged silence (exceeding 10 time steps) experience gradual opinion drift toward the perceived majority position, implementing the classical spiral of silence effect. Simultaneously, the model incorporates emotion-mediated influence dynamics, where opinion expression triggers both stance and emotion diffusion to neighboring agents, with influence magnitudes modulated by current emotion intensities, stance similarities, and global emotion distributions. This emotional contagion mechanism provides a computational pathway for the counter-spiral effects observed in empirical data.

In terms of interaction mechanics, we incorporated emotion-driven influence functions, where emotional intensity modulates both opinion transmission and emotional contagion across ties. For example, surprise and joy emotions—when prevalent—act as amplifiers of opinion shifts. The emotional intensity of an agent directly affects the strength of opinion influence on neighbors, with highly emotional agents exerting stronger influence regardless of their stance position. This creates asymmetric influence patterns that can generate non-linear opinion dynamics resembling those observed in empirical data.

To simulate the dynamic information environment characteristic of real-world technical controversies, we implemented randomized external influence events affecting subsets of the agent population, particularly during early simulation stages. External shocks (e.g., viral criticism events) and asymmetrical feedback loops are introduced to reflect real-world information disturbances. Notably, early steps include exogenous surges in "surprise" intensity to emulate unexpected revelations or scandals that commonly characterize technical controversy events. In our model, the intensity of emotional contagion is influenced by the global emotional distribution; for instance, when an emotion such as surprise becomes widespread in the network, its propagation rate is correspondingly enhanced, helping to explain the emotional clustering effects observed in empirical data.

The simulation was executed for 50 discrete time steps, providing sufficient temporal scope to observe both transient dynamics and emergent equilibrium states. Comprehensive data collection captured the complete state of each agent at every time step, enabling

detailed analysis of opinion trajectories, emotional diffusion patterns, and network structural evolution. This rich dataset allowed for both aggregate-level examination of opinion distribution dynamics and individual-level analysis of behavioral transitions, providing a computational laboratory for exploring the micro-mechanisms underlying observed macro-level discourse patterns. This model design enables us to examine how minority expressions (e.g., skepticism) can catalyze network-level mobilization and opinion reversals, providing computational support for the counter-spiral hypothesis introduced earlier.

4. Results

4.1. Descriptive Statistics and Temporal Evolution of Public Opinion

Analysis of the Manus controversy dataset reveals distinctive temporal patterns and user engagement dynamics throughout the five-day observation period. As shown in Figure 2, the controversy exhibited two prominent activity peaks: an initial surge on 6–7 March, followed by a substantially larger spike on 7 March, reaching approximately 2567 posts in a single hour. This dual-peak pattern is characteristic of technical controversies where initial product announcements generate moderate interest, followed by explosive engagement once technical scrutiny and detailed evaluations emerge. After the second peak, activity levels demonstrated a gradual decline with periodic minor fluctuations, suggesting sustained but diminishing interest in the controversy.

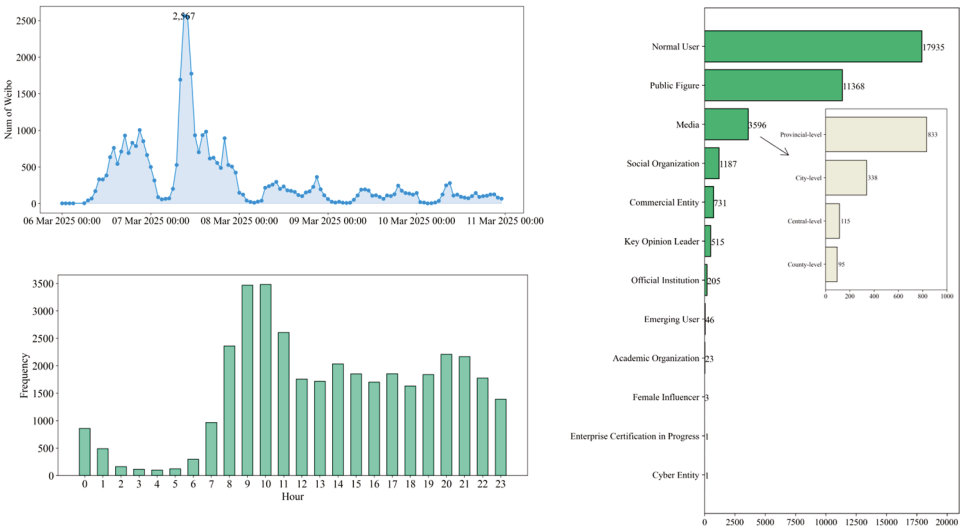


Figure 2. Trends in Manus controversy's microblogging data, cumulative interactions by time period, and user type structure distribution.

The hourly distribution analysis indicates that user engagement was most intensive during evening hours (8 PM to 10 PM), with peak activity occurring around 9 PM (3467 posts) and 10 PM (3480 posts). This temporal pattern aligns with typical social media usage behaviors in China, where evening hours represent prime engagement periods. Notably, even during low-activity hours (3 AM to 5 AM), the controversy maintained a baseline of approximately 100–150 posts per hour, indicating the sustained nature of public interest in this technical event.

Examination of user type distribution reveals that normal users constituted the predominant participant group (17,935 posts, 48.6%), followed by public figures (11,368 posts, 30.8%), and media accounts (3596 posts, 9.7%). Among media participants, provincial-level outlets demonstrated the highest engagement (833 posts), significantly outpacing both city-level media (338 posts) and central-level media (115 posts). This prominent role of provincial media suggests that regional information channels played a crucial function

in amplifying and framing the controversy, potentially reflecting regional interests in technological innovation narratives.

The emotional dynamics of the controversy, depicted in Figure 3, exhibit distinct patterns across different sentiment categories. Positive emotions, particularly joy and surprise, dominated the early stages of the controversy, with joy reaching near-maximum intensity (approximately 1.0) during the initial hours of 6 March. This was quickly followed by a spike in surprise sentiment, suggesting an initial phase of excitement and unexpected interest in the Manus announcement. Throughout the observation period, joy and surprise continued to oscillate between 0.3 and 0.6 intensity levels, maintaining primacy over other emotional expressions.

Figure 3. Results of sentiment categorization over time for the Manus controversy.

Negative emotions, including anger, disgust, fear, and sadness, remained comparatively subdued throughout most of the observation period, generally registering below 0.1 in intensity. However, notable spikes in anger (reaching 0.4) and sadness (reaching 0.5) occurred on 9 March, coinciding with the emergence of technical criticisms and revelations about potential code misappropriation. These emotional flares represent critical moments in the controversy's evolution, marking transitions in public sentiment as new information challenged initial positive perceptions.

The stance analysis presented in Figure 4 provides perhaps the most revealing insights into the controversy's evolution. Initial discourse was overwhelmingly positive, with bullish stances accounting for over 90% of expressions during the first hours of 6 March. This extremely high initial support gradually moderated but remained dominant (60–80%) throughout 6 March and early 7 March. However, a significant transition occurred during late 7 March and into 8 March, when skeptical stances began to gain substantial ground, increasing from under 20% to approximately 30–40% of expressions.

The most dramatic stance shift occurred on 8 March, when denial stances peaked at approximately 65% during a brief period, representing the controversy's inflection point. This coincided with the public revelation of technical evidence challenging Manus's originality claims. Following this critical juncture, stance distribution entered a more volatile phase, with bullish positions fluctuating between 40% and 80%, while skeptical and denial positions demonstrated responsive oscillation patterns.

Throughout the controversy's final days (9–10 March), stance distribution stabilized into a pattern of persistent contestation, with bullish stances maintaining a diminished but still dominant position (averaging 55–60%), counterbalanced by substantial skeptical stances (25–30%) and a smaller but consistent proportion of outright denial positions (10–15%). This final equilibrium state represents a significant moderation from the initial

near-unanimous positive stance, illustrating the evolution of public opinion as technical evidence and critical evaluation entered the discourse.

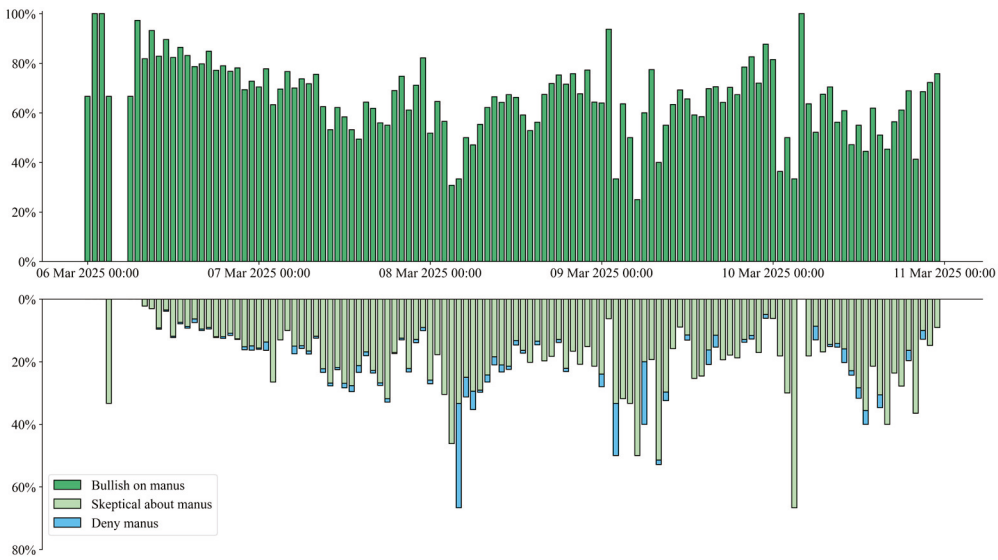


Figure 4. Results of stance categorization over time for the Manus controversy.

4.2. Lagged Interaction Effects of Stance

The temporal dynamics of stance positions in the Manus controversy reveal complex interaction patterns that extend beyond simple opinion distribution. As illustrated in Figure 5, the proportion of different stance positions exhibits distinctive fluctuation patterns throughout the observation period. The support for Manus (represented by the red line) dominated the early discourse with proportions exceeding 0.8 during the first day of the controversy. However, this initially strong positive position experienced several notable declines, particularly on 8 March when it temporarily dropped to approximately 0.45, coinciding with significant increases in both skeptical and neutral positions.

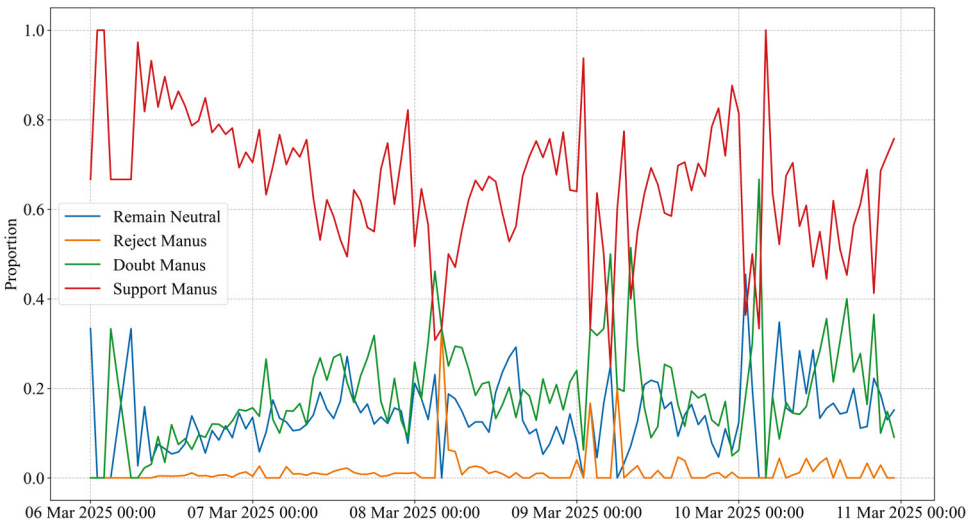


Figure 5. Changes in the proportion of different stances in the Manus controversy.

What emerges from this temporal analysis is a striking pattern of stance interactions characterized by rapid responsive shifts. Specifically, whenever doubt about Manus (green line) experiences a substantial increase, support for Manus typically demonstrates a corresponding surge shortly thereafter. This pattern is particularly evident during several

key moments: on 8 March, when doubt reached approximately 0.45, it was followed by a support rebound to 0.75; on 9 March, when a doubt peak reached 0.3, it was followed by a support surge to nearly 0.95; and again, on 10 March, multiple doubt increases were consistently followed by support rebounds.

This observation suggested the presence of systematic lagged interaction effects between opposing stance positions. To formally test this hypothesis, we conducted a cross-correlation function (CCF) analysis between the proportions of bullish (positive) and skeptical stance positions across the entire time series. Figure 6 displays the results of this analysis, with the horizontal axis representing the lag between the two time series (measured in 15-min intervals) and the vertical axis showing the correlation coefficient value.

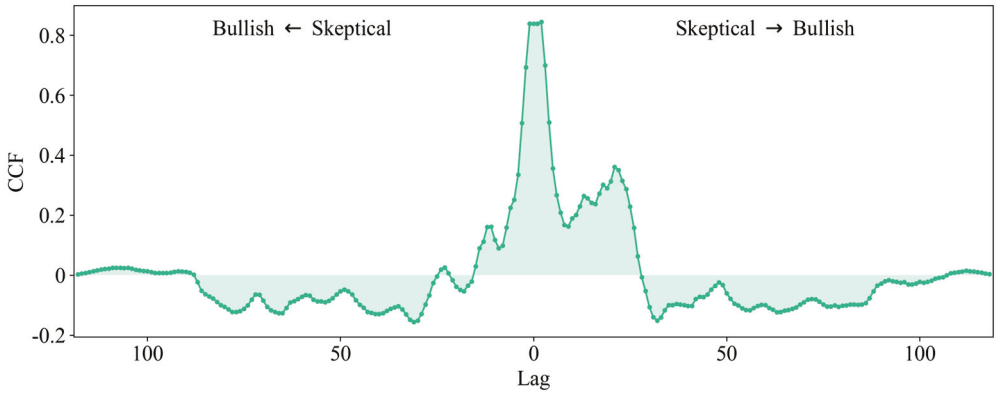


Figure 6. Cross-correlations between positive emotions (Bullish) and skeptical emotions (Skeptical) in the Manus controversy.

The CCF analysis reveals a striking asymmetric pattern. When skeptical stances lead bullish stances (positive lag values on the right side of the graph), we observe substantial positive correlation coefficients peaking at approximately 0.85 at around a 10–15 interval lag. This indicates that increases in skeptical sentiment are strongly associated with subsequent increases in bullish sentiment after approximately 2.5–3.75 h. This pattern suggests a consistent "counter-response" dynamic where expressions of doubt trigger defensive or supportive responses from those holding positive views about Manus.

Conversely, when bullish stances lead skeptical stances (negative lag values on the left side), we observe predominantly negative correlation coefficients, reaching approximately −0.15. This negative correlation indicates that increases in positive sentiment tend to be followed by decreases in skeptical sentiment, though this relationship is considerably weaker than the positive correlation in the opposite direction.

The asymmetric nature of these lagged correlations reveals a fundamental characteristic of stance dynamics in this technical controversy: skeptical positions appear to function as powerful triggers that stimulate responsive expression from those holding positive views, while positive expressions exert a milder suppressive effect on subsequent skeptical expressions. This asymmetry challenges simple conceptions of the spiral of silence, suggesting instead a complex "counter-spiral" phenomenon where perceived challenges to a dominant position mobilize defensive expression rather than inducing silence.

The strength of the positive correlation coefficient (reaching 0.85) indicates that this is not merely a statistical artifact but represents a robust pattern of interaction between opposing stance positions. The temporal specificity of the lag relationship—peaking at approximately 10–15 intervals—further suggests a systematic response dynamic rather than random fluctuation. This lagged interaction effect provides empirical evidence for the counter-spiral of silence phenomenon hypothesized in our research framework, demon-

strating how in digital technical discourse, expressions of minority or challenging views can trigger amplified expression from those holding majority or challenged positions.

4.3. Agent-Based Simulation of Sentiment and Stance Formation and Counter-Silence Dynamics

Building upon our empirical findings of lagged stance interactions, we implemented an agent-based simulation to explore the micro-level mechanisms that generate the observed counter-spiral phenomena. The simulation results, depicted in Figures 7 and 8, provide computational evidence for how individual agent behaviors and network structures evolve to produce the macro-level patterns identified in our empirical analysis.

Figure 7 illustrates the simulated evolution of stance proportions and emotional intensities over 50 time steps. The upper panel shows the progressive shift in the ratio between supporters and skeptics of Manus. Initially, the simulation was configured with approximately 45% skeptics and 55% supporters, reflecting the early distribution observed in the empirical data. As the simulation progressed, a notable transition occurred, with skeptic proportions gradually declining to approximately 38% by step 50, while supporter proportions increased to 62%. This trajectory mirrors the counter-spiral phenomenon observed in the actual controversy, where challenges to the dominant position ultimately triggered reinforcement rather than erosion of that position.

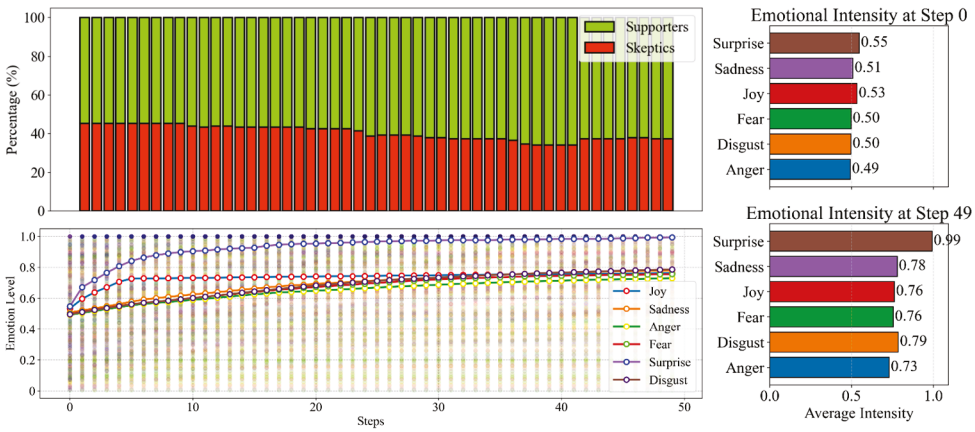


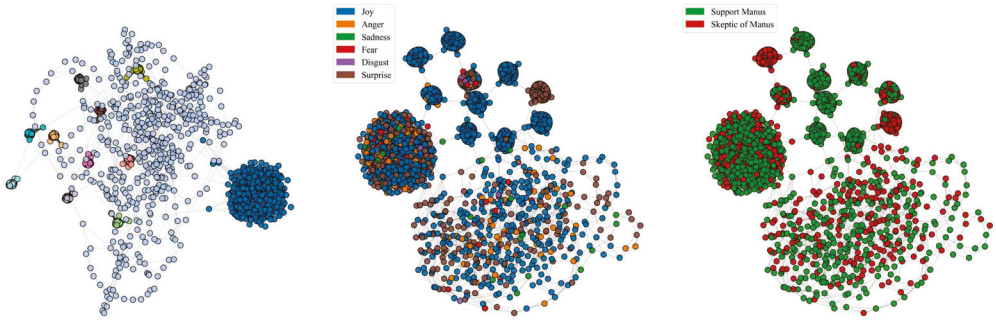
Figure 7. Evolution of stance and sentiment obtained by ABM modeling of the Manus controversy.

The emotional dynamics displayed in the lower panel of Figure 7 reveal how different emotions evolve throughout the simulation. Surprise emerges as the most intensified emotion, showing a rapid increase from an initial value of 0.55 to nearly 0.99 by step 15, after which it stabilizes at maximum intensity. This dominant surprise effect aligns with our empirical findings, where surprise played a crucial role in the discourse evolution. Joy demonstrates a more moderate yet steadily increasing trajectory, rising from 0.53 to 0.80 by the final step. The negative emotions—sadness, anger, fear, and disgust—all show gradual intensification from their initial values (ranging from 0.49 to 0.51) to final values between 0.73 and 0.87.

The comparison of emotional intensities between step 0 and step 49, displayed on the right side of Figure 7, quantifies these transformations. While all emotions experienced intensification, surprise showed the most dramatic increase (from 0.55 to 0.99), followed by disgust (0.50 to 0.87) and sadness (0.51 to 0.87). This differential emotional evolution suggests that controversy discourse generates asymmetric emotional responses, with surprise functioning as a primary driver of engagement and emotional contagion.

Figure 8 provides visualization of the network structures underlying the simulation, offering insights into how agent interactions shape opinion formation. The initial network structure (Figure 8a) displays the hybrid topology incorporating caveman, Barabási-Albert,

and Watts-Strogatz components, with nodes colored according to their dominant emotions. At this initial stage, the emotional distribution appears relatively random, with some clustering of joy (blue) and surprise (purple) evident in certain network regions.



(**a**) (**b**) (**c**)

Figure 8. Network visualization for ABM simulation of the Manus controversy. (**a**) Initial network structure. (**b**) Evolved network structure (by sentiment). (**c**) Evolved network structure (by stance).

As the simulation progresses to its final state, the network visualization by sentiment (Figure 8b) reveals significant emotional clustering and community formation. Nodes dominated by joy and surprise form distinct communities, with surprise-dominated nodes (purple) appearing particularly central in the network structure. This clustering effect demonstrates how emotional contagion operates through network connections, leading to localized emotional homogeneity and global emotional diversity.

The stance-based visualization of the evolved network (Figure 8c) provides the most direct evidence for the counter-spiral phenomenon. Supporters of Manus (green nodes) have formed several large, densely connected communities that dominate the network structure, while skeptics (red nodes) appear more fragmented and peripheral. This spatial distribution illustrates how the counter-spiral process manifests structurally, with supporters forming reinforcing communities that strengthen their collective position against skeptical challenges.

The expanded network size (1500 nodes total) allowed us to observe these phenomena with greater clarity and statistical reliability. The scale-free properties of the Barabási–Albert component are evident in the presence of highly connected hub nodes that facilitate rapid information spread, while the small-world characteristics of the Watts–Strogatz component enable efficient emotion propagation across otherwise distant network regions. The connected caveman structure creates tightly-knit communities that reinforce shared stances and emotional states, contributing to the formation of opinion enclaves resistant to outside influence.

The simulation results provide computational validation for the lagged interaction effects observed in our empirical analysis. The agent-based model successfully reproduces several key phenomena: (1) the gradual increase in supporter proportions following initial skeptical challenge; (2) the differential evolution of emotional intensities, with surprise emerging as dominant; (3) the formation of emotion-based and stance-based communities within the network structure; and (4) the structural advantage gained by the supporter position through more effective community formation.

These findings suggest that the counter-spiral phenomenon emerges from the interaction of several mechanisms: emotional contagion, where emotions—particularly surprise—spread rapidly through network connections; preferential attachment, where agents with similar positions tend to form denser connection patterns; and counter-mobilization, where challenges to a position trigger defensive expression rather than silence. The simu-

lation thus provides a computational mechanism for understanding how the lagged stance interactions observed in the empirical data might emerge from individual agent behaviors operating within complex network structures.

5. Discussion

The empirical findings and simulation results from our analysis of the Manus controversy offer significant insights into the dynamics of stance formation and opinion reversal in contemporary technical discourse environments. Our observation of lagged stance interactions and counter-spiral phenomena invites deeper theoretical reflection on how digital communication environments reconfigure traditional public opinion processes.

5.1. Lagged Stance Interactions as Dynamic Catalysts

The cross-correlation analysis revealing strong positive associations between skeptical expressions and subsequent bullish responses (r = 0.85 at 10–15 interval lag) illuminates a fundamental characteristic of technical discourse dynamics that challenges conventional understanding of opinion formation. This pronounced lagged interaction effect demonstrates that, contrary to traditional conceptions of minority opinion suppression, skeptical challenges in technical contexts appear to function as catalytic triggers that mobilize defensive responses from supporters. The asymmetric nature of this relationship—with skeptical expressions more powerfully influencing subsequent bullish expressions than the reverse—suggests a fundamentally reactive pattern of discourse engagement where perceived threats to dominant technical narratives generate disproportionate counter-mobilization.

This finding aligns with but extends recent theoretical work on "reactance theory" in digital environments, where psychological reactance to perceived opinion threats generates amplified counter-responses rather than opinion moderation. In the Manus case, the temporal specificity of this lag relationship (peaking at approximately 2.5–3.75 h) provides particularly valuable insight into the operational timescale of digital counter-response dynamics, offering a potential "response window" during which strategic communication interventions might be most effective.

5.2. Counter-Spiral of Silence in Technical Discourse

Before analyzing how the Manus controversy deviates from classical spiral dynamics, it is essential to reiterate the conceptual logic of the original model. Noelle-Neumann theorized that individuals continuously scan their environment to assess opinion climates. When perceiving themselves to be in the minority, they anticipate social exclusion, leading to self-silencing. Over time, this silencing reinforces the dominance of the perceived majority, creating a spiral. In our study, however, we observe the opposite dynamic: expressions of doubt do not disappear, but instead provoke intensified, emotionally driven responses from supportive actors. This suggests a reversal of the spiral logic—what we define as a "counter-spiral"—where minority expressions act as triggers, not inhibitors, of further discourse. The mechanisms driving this reversal appear rooted in digital affordances (e.g., low-cost rebuttals, real-time feedback) and the emotional architecture of controversy, particularly the mobilizing role of surprise and indignation. Thus, the Manus case reflects a structurally inverted version of the spiral: a reactive, nonlinear escalation rather than a suppression spiral.

The observed counter-spiral phenomenon—where challenges to the dominant position ultimately trigger reinforcement rather than erosion of that position—fundamentally contradicts classical Spiral of Silence predictions. In traditional Spiral of Silence theory, the perception of one's opinion as a minority should lead to self-censorship and silence. Yet,

our agent-based simulations successfully reproduced empirical patterns showing how supporter communities formed increasingly dense network structures in response to skeptical challenges, ultimately strengthening rather than weakening their collective position.

This counterintuitive finding can be explained through several mechanisms illuminated by our simulation results. First, emotional contagion—particularly surprise—appears to function as a potent amplifier of stance expression, spreading rapidly through network connections and intensifying engagement with contested positions. Second, homophilous clustering generates protected discursive spaces where like-minded individuals reinforce shared positions, creating what might be termed "opinion sanctuaries" resistant to external pressure. Third, the multi-channel nature of digital communication environments enables simultaneous expression across diverse platforms, reducing the fear of isolation that underpins traditional silencing mechanisms.

5.3. Technological Nationalism and Cognitive Asymmetries

The Manus controversy exhibits characteristics of broader technological nationalism processes, revealing fundamental cognitive asymmetries in how diverse stakeholders interpret technological innovations. The controversy's essence reflects a significant cognitive gap in the AI democratization process: public expectations aligned with "ChatGPT-like disruption" colliding with an industry reality of incremental engineering optimization. The opinion reversal trajectory (6 March viral emergence → 7 March account suspension → 13 March code leak) reveals two critical patterns underlying technological discourse dynamics.

First, technical knowledge asymmetry intensifies emotional polarization. Our data demonstrates how ordinary users exhibited romantic conceptualizations of technical terminology such as "multi-agent architecture", while developer communities focused on practical constraints such as token costs and API dependencies. This discrepancy represents a form of "technological expectation gap" that generates emotional volatility when public imagination confronts technical limitations. The emotional intensification observed in our simulation results—where surprise emerged as the dominant emotion, increasing from 0.55 to 0.99—reflects this collision between idealized technological expectations and emerging technical realities.

Second, digital platforms' inherent affordances accelerate technological demystification. The "emotional resonance-stance reversal" mechanism documented in our analysis enabled technical flaws to complete a full "mythmaking–questioning–rationalization" cycle within 48 h. This accelerated demystification process represents a distinctive feature of contemporary technical discourse environments, where the traditional vetting mechanisms of scientific and technical communities have been partially supplanted by rapid, emotionally-driven public evaluation cycles.

These findings carry significant implications for both theory development and practical communication strategies in technical contexts. For theoretical advancement, our results suggest the need for a fundamental reconsideration of silence dynamics in digital environments, particularly regarding how network structure, emotional contagion, and technical knowledge disparities modulate traditional opinion expression processes. For practical application, our findings highlight the critical importance of managing technical expectation gaps, anticipating counter-response dynamics, and developing communication strategies attentive to the specific temporal windows during which stance interactions are most volatile.

The Manus case ultimately serves as a cautionary exemplar for the AI industry, highlighting how the dual pressures of "technological breakthrough anxiety" and "capital monetization imperatives" create precarious communication environments. Successful navi-

gation of these environments requires a strategic balance between marketing volume and technical transparency, avoiding the increasingly common pattern where viral popularity immediately precipitates critical scrutiny and potential reputational collapse.

6. Conclusions

6.1. Theoretical Contributions

This research makes several significant theoretical contributions to understanding public opinion dynamics in digital environments. First, we have empirically documented and conceptually developed the "Counter-Spiral of Silence" phenomenon, demonstrating how digital communication environments can reverse traditional silencing mechanisms through interactions of emotional contagion, network structure, and stance dynamics. By identifying the robust positive cross-correlation between skeptical expressions and subsequent bullish responses, we provide quantitative evidence challenging classical assumptions about opinion minority behaviors.

Second, our integration of agent-based modeling with empirical stance analysis advances methodological approaches to studying opinion dynamics. The successful reproduction of observed counter-spiral patterns through computational simulation demonstrates how individual-level behavioral rules generate emergent macro-level discourse patterns that traditional analytical approaches struggle to explain.

Third, our research extends traditional spiral of silence theory by incorporating temporal dynamics and emotional contagion mechanisms specific to technical discourse environments. By documenting how emotions (particularly surprise) spread through network connections and modulate stance expression, we provide a more nuanced understanding of the affective dimensions underlying public opinion formation.

6.2. Practical Implications

Our findings offer valuable practical insights for technology developers, communication strategists, and platform governance. For technology companies, the documented lag period between skeptical challenges and supportive mobilization (approximately 2.5–3.75 h) provides a specific operational window during which strategic interventions might effectively address emerging criticisms before counter-mobilization intensifies discourse polarization.

For communication strategists, our identification of surprise as the dominant emotion in technical discourse offers guidance for message framing. The simulation results showing how surprise rapidly propagates through networks suggest that controlling narrative surprise elements may be crucial for managing discourse evolution. Additionally, the documented network clustering effects highlight the importance of engaging with diverse community structures rather than treating public opinion as homogeneous.

For platform governance, our research demonstrates how technical discourse evolves through distinctive patterns different from other forms of public opinion formation. The rapid progression from near-unanimous support to contested stance distribution within a compressed timeframe suggests potential benefits of implementing "cooling off" mechanisms or specialized moderation approaches for technically complex topics.

6.3. Limitations and Future Directions

While our research provides valuable insights into lagged stance interactions and counter-spiral phenomena, several limitations should be acknowledged. Our analysis focused on a single case study (Manus) within a specific platform (Weibo) and cultural context, limiting generalizability to other technologies or communication environments. The temporal scope was also constrained to the initial discourse formation period (five days),

potentially missing longer-term evolutionary patterns. Additionally, our agent-based model, while improved, still employs simplified representations of user characteristics that may not fully capture the complexity of real-world discourse participation.

Based on these limitations, we propose the following specific research directions for future investigation:

(1) Cross-platform comparative research: Future studies should extend similar methodological approaches across multiple platforms (e.g., Zhihu, Bilibili, Twitter, Reddit) to examine how different platform affordances, moderation policies, and user demographics influence counter-spiral dynamics. Such comparative analysis could identify platform-specific factors that either amplify or attenuate counter-mobilization responses.

(2) Extended longitudinal analysis: Expanding the temporal scope to 30–60 days would enable investigation of how counter-spiral patterns evolve beyond initial controversy peaks, potentially revealing cyclical patterns, delayed effects, or eventual equilibrium states that our compressed observation period could not capture.

(3) Cultural and linguistic variation studies: Applying our computational framework to technical controversies across different cultural contexts (e.g., Western, East Asian, and Global South settings) would illuminate how technological nationalism and expectation gaps manifest differently across various innovation ecosystems and communication norms.

(4) Multi-method validation: Combining computational approaches with qualitative interviews of social media users involved in technical controversies would provide deeper insights into the psychological mechanisms underlying stance expression and emotional responses that cannot be fully captured through digital trace data alone.

(5) Intervention design and testing: Developing and evaluating specific communication strategies during the critical response window (2.5–3.75 h) identified in our study would offer practical guidance for technology communicators seeking to moderate polarization during emerging controversies.

These research directions would address the limitations of our current study while extending both theoretical understanding and practical approaches to navigating the increasingly complex landscape of public engagement with technological innovation.

Author Contributions: Conceptualization, K.Z. and C.D.; methodology, K.Z. and W.Z.; software, W.Z. and Y.G.; validation, K.Z., C.D. and G.Y.; formal analysis, K.Z. and W.Z.; investigation, C.D.; re-sources, Y.G.; data curation, C.D. and W.Z.; writing—original draft preparation, K.Z., C.D. and W.Z.; writing—review and editing, G.Y. and Y.G.; visualization, K.Z.; supervision, G.Y., Y.G. and J.M.; project administration, J.M.; funding acquisition, Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant No. 72374056 and "HIT-China Mobile" 5G Application Innovation Joint Research Institute Project under Grant No. 2023B012.

Data Availability Statement: This article encompasses this study's original contributions; further inquiries can be addressed to the corresponding author.

Acknowledgments: The authors would like to thank the National Natural Science Foundation of China and "HIT-China Mobile" 5G Application Innovation Joint Research Institute for providing the funding to support this research. They would also like to thank Zhiwei Research Institute for providing the data collection platform.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cinelli, M.; Quattrociocchi, W.; Galeazzi, A.; Valensise, C.M.; Brugnoli, E.; Schmidt, A.L.; Zola, P.; Zollo, F.; Scala, A. The COVID-19 Social Media Infodemic. *Sci. Rep.* **2020**, *10*, 16598. [CrossRef] [PubMed]
2. Vosoughi, S.; Roy, D.; Aral, S. The Spread of True and False News Online. *Science* **2018**, *359*, 1146–1151. [CrossRef] [PubMed]
3. Pennycook, G.; Rand, D.G. The Psychology of Fake News. *Trends Cogn. Sci.* **2021**, *25*, 388–402. [CrossRef] [PubMed]
4. Perrin, A.J.; McFarland, K. Social theory and public opinion. *Annu. Rev. Sociol.* **2011**, *37*, 87–107. [CrossRef]
5. Noelle-Neumann, E. The Spiral of Silence: A Theory of Public Opinion. *J. Commun.* **1974**, *24*, 43–51. [CrossRef]
6. Neubaum, G.; Krämer, N.C. Monitoring the opinion of the crowd: Psychological mechanisms underlying public opinion perceptions on social media. *Media Psychol.* **2017**, *20*, 502–531. [CrossRef]
7. Fox, J.; Holt, L.F. Fear of Isolation and Perceived Affordances: The Spiral of Silence on Social Networking Sites Regarding Police Discrimination. *Mass Commun. Soc.* **2018**, *21*, 533–554. [CrossRef]
8. Entman, R.M.; Usher, N. Framing in a fractured democracy: Impacts of digital technology on ideology, power and cascading network activation. *J. Commun.* **2018**, *68*, 298–308. [CrossRef]
9. Gao, C.; Lan, X.; Li, N.; Yuan, Y.; Ding, J.; Zhou, Z.; Xu, F.; Li, Y. Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 1259. [CrossRef]
10. Wu, Z.; Peng, R.; Han, X.; Zheng, S.; Zhang, Y.; Xiao, C. Smart Agent-Based Modeling: On the Use of Large Language Models in Computer Simulations. *arXiv* **2023**, arXiv:2311.06330.
11. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the NAACL—HLT, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
12. Sun, C.; Qiu, X.; Xu, Y.; Huang, X. How to Fine—Tune BERT for Text Classification? In *China National Conference on Chinese Computational Linguistics*; Springer: Kunming, China, 2019; pp. 194–206.
13. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models Are Few—Shot Learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
14. Squazzoni, F.; Jager, W.; Edmonds, B. Social Simulation in the Social Sciences: A Brief Overview. *Soc. Sci. Comput. Rev.* **2014**, *32*, 279–294. [CrossRef]
15. Macy, M.W.; Willer, R. From Factors to Actors: Computational Sociology and Agent—Based Modeling. *Annu. Rev. Sociol.* **2002**, *28*, 143–166. [CrossRef]
16. Starbird, K.; Arif, A.; Wilson, T. Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations. *Proc. ACM Hum. Comput. Interact.* **2019**, *3*, 1–26. [CrossRef]
17. Sigala, M. Social Media and Crisis Management in Tourism: Applications and Implications for Research. *Inf. Technol. Tour.* **2011**, *13*, 269–283. [CrossRef]
18. Lazer, D.M.J.; Baum, M.A.; Benkler, Y.; Berinsky, A.J.; Greenhill, K.M.; Menczer, F.; Metzger, M.J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; et al. The Science of Fake News. *Science* **2018**, *359*, 1094–1096. [CrossRef] [PubMed]
19. Scheufele, D.A.; Moy, P. Twenty-Five Years of the Spiral of Silence: A Conceptual Review and Empirical Outlook. *Int. J. Public Opin. Res.* **2000**, *12*, 3–28. [CrossRef]
20. Glynn, C.J.; Hayes, A.F.; Shanahan, J. Perceived Support for One's Opinions and Willingness to Speak Out: A Meta-Analysis of Survey Studies on the "Spiral of Silence". *Public Opin. Q.* **1997**, *61*, 452–463. [CrossRef]
21. Kahl, S.J.; Grodal, S. Discursive Strategies and Radical Technological Change: Multilevel Discourse Analysis of the Early Computer (1947–1958). *Strateg. Manag. J.* **2016**, *37*, 149–166. [CrossRef]
22. Valkenburg, P.M.; Peter, J.; Walther, J.B. Media Effects: Theory and Research. *Annu. Rev. Psychol.* **2016**, *67*, 315–338. [CrossRef]
23. Matthes, J.; Hayes, A.F.; Rojas, H.; Shen, F.; Min, S.J.; Dylko, I.B. Exemplifying a Dispositional Approach to Cross-Cultural Spiral of Silence Research: Fear of Social Isolation and the Inclination to Self-Censor. *Int. J. Public Opin. Res.* **2012**, *24*, 287–305. [CrossRef]
24. Stoycheff, E. Under Surveillance: Examining Facebook's Spiral of Silence Effects in the Wake of NSA Internet Monitoring. *J. Mass Commun. Q.* **2016**, *93*, 296–311. [CrossRef]
25. Brady, W.J.; Crockett, M.J.; Van Bavel, J.J. The MAD model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspect. Psychol. Sci.* **2020**, *15*, 978–1010. [CrossRef] [PubMed]
26. Donovan, J.; Boyd, D. Stop the presses? Moving from strategic silence to strategic amplification in a networked media ecosystem. *Am. Behav. Sci.* **2021**, *65*, 333–350. [CrossRef]
27. Luarn, P.; Hsieh, A.Y. Speech or Silence: The Effect of User Anonymity and Member Familiarity on the Willingness to Express Opinions in Online Communities. *Online Inf. Rev.* **2014**, *38*, 881–895. [CrossRef]
28. Wojcieszak, M.; Price, V. What Underlies the False Consensus Effect? How Personal Opinion and Disagreement Affect Perception of Public Opinion. *Int. J. Public Opin. Res.* **2009**, *21*, 25–46. [CrossRef]
29. Willaert, T.; Van Eecke, P.; Beuls, K.; Steels, L. Building social media observatories for monitoring online opinion dynamics. *Soc. Media+ Soc.* **2020**, *6*, 2056305119898778. [CrossRef]

30. Mohammad, S.M.; Sobhani, P.; Kiritchenko, S. Stance and Sentiment in Tweets. *ACM Trans. Internet Technol.* **2017**, *17*, 1–23. [CrossRef]
31. Küçük, D.; Can, F. Stance Detection: A Survey. *ACM Comput. Surv.* **2020**, *53*, 1–37. [CrossRef]
32. Fan, L.; Li, L.; Ma, Z.; Lee, S.; Yu, H.; Hemphill, L. A bibliometric review of large language models research from 2017 to 2023. *ACM Trans. Intell. Syst. Technol.* **2024**, *15*, 1–25. [CrossRef]
33. Wankhade, M.; Rao, A.C.S.; Kulkarni, C. A Survey on Sentiment Analysis Methods, Applications, and Challenges. *Artif. Intell. Rev.* **2022**, *55*, 5731–5780. [CrossRef]
34. Mohammad, S.M.; Kiritchenko, S.; Sobhani, P.; Zhu, X.; Cherry, C. SemEval-2016 Task 6: Detecting Stance in Tweets. In Proceedings of the 10th International Workshop on Semantic Evaluation, San Diego, CA, USA, 16–17 June 2016; pp. 31–41.
35. Li, Y.; Caragea, C.; Caragea, D.; Karami, A. Improving Stance Detection with Multi-Dataset Learning and Knowledge Distillation. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 10655–10667.
36. Kawintiranon, K.; Singh, L. Knowledge Enhanced Masked Language Model for Stance Detection. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics, Online, 6–11 June 2021; pp. 4725–4735.
37. Beltagy, I.; Lo, K.; Cohan, A. SciBERT: A Pretrained Language Model for Scientific Text. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, Hong Kong, China, 3–7 November 2019; pp. 3615–3620.
38. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Yang, Z.; Wang, S.; Hu, G. Pre-Training with Whole Word Masking for Chinese BERT. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 3504–3514. [CrossRef]
39. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.; Le, Q.; Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 24824–24837.
40. Yang, Z.; Dong, L.; Du, X.; Cheng, H.; Cambria, E.; Liu, X.; Gao, J.; Wei, F. Language models as inductive reasoners. *arXiv* **2022**, arXiv:2212.10923.
41. Zhong, R.; Lee, K.; Zhang, Z.; Klein, D. Adapting Language Models for Zero-Shot Learning by Meta-Tuning on Dataset and Prompt Collections. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2021, Punta Cana, Dominican Republic, 7–11 November 2021; pp. 2856–2878.
42. Gururangan, S.; Marasović, A.; Swayamdipta, S.; Lo, K.; Beltagy, I.; Downey, D.; Smith, N.A. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 8342–8360.
43. Stahl, G.K.; Filatotchev, I.; Ireland, R.D.; Miska, C. Five decades of research on the role of context in management: From universalism toward contingent, multilevel and polycontextual perspectives. *Acad. Manag. Collect.* **2023**, *2*, 1–18. [CrossRef]
44. Lai, M.; Cignarella, A.T.; Farías, D.I.; Bosco, C.; Patti, V.; Rosso, P. Multilingual Stance Detection in Social Media Political Debates. *Comput. Speech Lang.* **2021**, *65*, 101075. [CrossRef]
45. Burton, J.W.; Lopez-Lopez, E.; Hechtlinger, S.; Rahwan, Z.; Aeschbach, S.; Bakker, M.A.; Hertwig, R. How Large Language Models Can Reshape Collective Intelligence. *Nat. Hum. Behav.* **2024**, *8*, 1643–1655. [CrossRef] [PubMed]
46. Flache, A.; Mäs, M.; Feliciani, T.; Chattoe-Brown, E.; Deffuant, G.; Huet, S.; Lorenz, J. Models of Social Influence: Towards the Next Frontiers. *J. Artif. Soc. Soc. Simul.* **2017**, *20*, 2. [CrossRef]
47. Piedrahita, P.; Borge-Holthoefer, J.; Moreno, Y.; González-Bailón, S. The Contagion Effects of Repeated Activation in Social Networks. *Soc. Netw.* **2018**, *54*, 326–335. [CrossRef]
48. Keijzer, M.A.; Mäs, M.; Flache, A. Communication in Online Social Networks Fosters Cultural Isolation. *Complexity* **2018**, *2018*, 9502872. [CrossRef]
49. Epstein, J.M. Agent-Based Computational Models and Generative Social Science. *Complexity* **1999**, *4*, 41–60. [CrossRef]
50. Sohn, D.; Geidner, N. Collective Dynamics of the Spiral of Silence: The Role of Ego-Network Size. *Int. J. Public Opin. Res.* **2016**, *28*, 25–45. [CrossRef]
51. Xiong, F.; Liu, Y. Opinion Formation on Social Media: An Empirical Approach. *Chaos Interdiscip. J. Nonlinear Sci.* **2014**, *24*, 013130. [CrossRef]
52. Stefanelli, A.; Seidl, R. Moderate and Polarized Opinions: Using Empirical Data for an Agent-Based Model. In Proceedings of the 16th Conference of the European Social Simulation Association, Cracow, Poland, 20–24 September 2021; pp. 1–12.
53. Acerbi, A. A cultural evolution approach to digital media. *Front. Hum. Neurosci.* **2016**, *10*, 636. [CrossRef]
54. Hassan, S.; Arroyo, J.; Galán, J.M.; Antunes, L.; Pavón, J. Asking the Oracle: Introducing Forecasting Principles into Agent-Based Modelling. *J. Artif. Soc. Soc. Simul.* **2013**, *16*, 13. [CrossRef]
55. Edmonds, B.; Le Page, C.; Bithell, M.; Chattoe-Brown, E.; Grimm, V.; Meyer, R.; Montañola-Sales, C.; Ormerod, P.; Root, H.; Squazzoni, F. Different Modelling Purposes. *J. Artif. Soc. Soc. Simul.* **2019**, *22*, 6. [CrossRef]
56. Bail, C.A.; Argyle, L.P.; Brown, T.W.; Bumpus, J.P.; Chen, H.; Hunzaker, M.F.; Lee, J.; Mann, M.; Merhout, F.; Volfovsky, A. Exposure to Opposing Views on Social Media Can Increase Political Polarization. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 9216–9221. [CrossRef] [PubMed]

57. De Domenico, M.; Granell, C.; Porter, M.A.; Arenas, A. The Physics of Spreading Processes in Multilayer Networks. *Nat. Phys.* **2016**, *12*, 901–906. [CrossRef]
58. Dehkordi, M.A.E.; Lechner, J.; Ghorbani, A.; Nikolic, I.; Chappin, E.; Herder, P. Using machine learning for agent specifications in agent-based models and simulations: A critical review and guidelines. *J. Artif. Soc. Soc. Simul.* **2023**, *26*, 9. [CrossRef]
59. Belkin, M. Fit without fear: Remarkable mathematical phenomena of deep learning through the prism of interpolation. *Acta Numer.* **2021**, *30*, 203–248. [CrossRef]
60. Grimm, V.; Berger, U.; Calabrese, J.M.; Cortés-Avizanda, A.; Ferrer, J.; Franz, M.; Groeneveld, J.; Hartig, F.; Jakoby, O.; Jovani, R.; et al. Using the ODD protocol and NetLogo to replicate agent-based models. *Ecol. Model.* **2025**, *501*, 110967. [CrossRef]
61. Zhang, B.; Dai, G.; Niu, F.; Yin, N.; Fan, X.; Wang, S.; Cao, X.; Huang, H. A Survey of Stance Detection on Social Media: New Directions and Perspectives. *arXiv* **2024**, arXiv:2409.15690.
62. Sohn, D. Spiral of silence in the social media era: A simulation approach to the interplay between social networks and mass media. *Commun. Res.* **2022**, *49*, 139–166. [CrossRef]
63. Laor, T. Breaking the silence: The role of social media in fostering community and challenging the spiral of silence. *Online Inf. Rev.* **2024**, *48*, 710–724. [CrossRef]
64. Kuncoro, H.R.; Hasanah, K. How Social Media Algorithms Potentially Reinforce Radical Views. *Insign. J. Int. Relat.* **2024**, *11*, 126–149. [CrossRef]
65. Chen, Z. An agent-based model for information diffusion over online social networks. *Pap. Appl. Geogr.* **2019**, *5*, 77–97. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

systems

MDPI

Article

Effect of Network Structure on Conflict and Project Value Creation

Cong Liu [1], Yuan Shan [2,3,*] and Jiming Cao [2]

[1] School of Economics and Management, Shanghai Polytechnic University, Shanghai 201209, China; liucong@sspu.edu.cn
[2] School of Economics and Management, Tongji University, Shanghai 200092, China; caojm@tongji.edu.cn
[3] Faculty of Economics and Business Administration, Ghent University, 9000 Ghent, Belgium
* Correspondence: 2010353@tongji.edu.cn

Abstract: This study explored the impact of network structure on conflict and project value creation. Network density and network centrality are two network structure dimensions. A survey was undertaken among professionals working in Chinese construction projects. A total of 308 surveys were analyzed using the structural equation model. The results revealed that network centrality has a negative impact on project value creation while network density has a positive impact. Network centrality has a negative impact on substantive conflicts but a positive impact on affective conflicts. The link between centrality and project value creation is weakened by substantive conflict but strengthened by affective conflict. This research gives a new direction for construction project governance and project value management. Furthermore, this research validates the constructive role of substantive conflicts, as well as the destructive impact of affective conflicts, thereby adding to the literature on conflict governance.

Keywords: network structure; conflict; project value creation; project governance

1. Introduction

With the increasing demand for infrastructure over the previous two decades, the quantity and scale of construction projects have increased [1]. Construction projects currently include the following characteristics: high investment, extended construction cycle, high complexity, dynamic environment, and multiple participating organizations (e.g., general contractors, subcontractors, design units, survey units, supervision units, suppliers) [2]. These organizations involved in the construction project vary in their professional expertise, skills, and capacities. No organization possesses all of the skills required to finish construction projects [3]. As a result, organizations build collaborative ties and a project network to complete construction projects. The project network is a vital route via which organizations obtain information, expertise, and resources [4]. However, different network structures (e.g., high-density network, high-centrality network) may have varied effects on construction project implementation. The structure of a project network, according to social capital theory, impacts whether and to what extent organizations embedded in the network gain resources (e.g., information, expertise, experience) [5]. Furthermore, the structure of the project network dictates how organizations incorporated in the network communicate with one another [6]. This may have an impact on organizational communication and cooperation, influencing the completion of construction project tasks and project value creation.

Construction projects are one-time, complex, uncertain, and characterized by incomplete contracts. Organizations engaged in the construction project network have varying levels of expertise, technology, and interest demands [7]. Therefore, organizational conflicts are common. In construction projects, interorganizational conflicts include substantive conflicts and affective conflicts. Affective conflict reflects organizational incompatibility, manifested as negative emotions between organizations, as well as explicit conflicting behaviors such as negative confrontation that arise when the organization feels uncoordinated or inconsistent with other organizations. Substantive conflict refers to different perspectives of an organization regarding the content of a project task, as well as different perspectives, contradictions, or disagreements of the organization regarding the specific process arrangements (e.g., schedule arrangement, project plan formulation, and resource allocation arrangement). If conflicts between organizations are not appropriately addressed or resolved in a timely manner, they may impede cooperation and the fulfillment of project duties [8]. Finally, this may have an impact on project value creation. Project value creation refers to achieving the goals of the project (i.e., quality, schedule, cost, and safety) and adding value to the owners and users of the construction project. According to social capital theory, the network structure (i.e., network density and network centrality) of construction projects may have an impact on interorganizational conflicts [5]. The high-centrality network contains organizations with significant power and authority. They may have control over the flow of resources (e.g., information, expertise, and experience) in the network [9]. However, this may distort information transfer and impede collaboration between organizations. This may lead to interorganizational disagreements during project implementation, resulting in interorganizational affective conflicts (i.e., interorganizational incompatibility manifested as negative emotions, as well as hostile behavior such as negative confrontation), which negatively affect project implementation and project value creation.

A high-density network is a highly interconnected communication network that promotes communication between organizations and increases the flexibility of their interactions, thus improving the efficiency of information dissemination and achieving task implementation coordination more quickly [10]. Simultaneously, a high-density network is typically more active, which aids in the sharing of knowledge and information between organizations, reduces interorganizational affective conflicts, and thus promotes the improvement of cooperation efficiency and effectiveness between organizations and project value creation [11]. The current research on network structure is mostly focused on soft elements in construction projects such as trust, communication, and leadership [11–14]. Few studies have been conducted to investigate whether network structure influences interorganizational conflicts and project value creation. This study investigates the impact of network structure on interorganizational conflicts and project value creation. This research splits network structure into two dimensions: network centrality and network density, and widens interorganizational conflicts into two types: substantive conflicts and affective conflicts. The research findings aid the project management team in better detecting the impact of network structure and in designing efficient project management strategies, ultimately promoting project value creation.

2. Literature Review

2.1. Network Structure

Over the last decade, social capital theory has been widely used to describe the relationship between organizational network structure (such as a trade network or a communication network) and organizational output (such as organizational performance) [15]. According to social capital theory, social capital is the whole of an organization's actual

and potential resources that are owned, available, and generated from its social interaction network [16]. It is widely assumed that social capital has three dimensions: (1) a structural dimension, which refers to the overall structure of the organization's social relationship; (2) a relationship dimension, which refers to mutual respect and trust developed during the interaction between organizations; and (3) a cognitive dimension, which refers to the implicit resources that help organizations produce shared norms or collective goals [17]. Even with clarification of these three dimensions, social capital is a complex sociological structure, making it difficult to quantitatively study.

In the context of construction management, social capital theory has been applied in various aspects. Regarding project collaboration, social capital plays a pivotal role in facilitating cooperation among diverse stakeholders [15]. For instance, high levels of trust (a key element of relational social capital) between owners and contractors can lead to more open communication, reduced opportunistic behavior, and lower transaction costs. A study by Cheraghi et al. (2023) demonstrated that in construction projects where there was a strong foundation of trust among participants, disputes were resolved more amicably, and the overall project delivery was more efficient [2]. In terms of information sharing, the network structure provided by structural social capital enables the rapid dissemination of information [15]. Subcontractors with well-connected networks are more likely to access valuable technical knowledge and market information, which can enhance their operational capabilities and contribute to the project's success. When it comes to resource integration, social capital allows construction firms to access external resources beyond their own organizational boundaries [9]. Through their social networks, firms can establish partnerships with research institutions to gain access to innovative construction technologies (cognitive social capital in the form of shared knowledge goals). This integration of resources not only improves the firm's competitiveness but also has a positive impact on the overall project quality and sustainability.

The majority of extant network research focuses on assessing the structural dimension of social capital (such as network structure). Taking into account the characteristics of construction project networks [4–7], this study also focuses on the structural dimension of social capital to investigate the relationship between construction project network structure, project conflict, and project value creation. In reference to previous research on network structure, the two representative indicators of network centrality and network density were selected to reflect network structure [9,11,13]. Network density is one of the project network's structural characteristic indicators [18]. Network density is commonly used to represent the closeness of relationships between stakeholders. In a construction project, stakeholders are linked and interact through cooperative relationships [5]. These collaboration links will progressively coalesce into a complex project network, influencing how organizations access resources and information [4]. The network density in the construction project network can be used to examine the degree of resource dissemination and measure the degree of contact of stakeholders in the project network [19]. Because of the homogeneity of stakeholders, it is advantageous for various stakeholders to create common aims and values through communication and cooperation. According to Zeng et al. (2022), high-density networks assist network members engage in more frequent activities [20]. Cao et al. (2017) proposed that increasing network density can better coordinate the behavior of network members [21]. Tang et al. (2019) argues that high-density communication networks aid in the achievement of organizational goals [22]. Therefore, increasing project network density may enhance the creation of trust between organizations and encourage stakeholders to access more information.

The degree to which an organization is placed at the core of the network is characterized as network centrality [23]. Network centrality is frequently used to express

network participants' status, influence, and reputation. High network centrality suggests that some actors are in the network's focal point, controlling and coordinating information transmission and resource allocation [24]. Network centrality is typically divided into three categories: degree centrality, close centrality, and between centrality. Specifically, if a network actor is directly related to many other network members, the network actor has a high degree of degree centrality. If a network actor can connect to several other network actors via a short path, the network actor's close centrality is high. Between centrality refers to the shortest path for one network actor relative to other network actors [25]. According to the three definitions of network centrality, network centrality primarily evaluates a network actor's ability to access resources, control other actors, and avoid being controlled by other actors [26]. Because construction project resources are limited, highly centralized stakeholders play a controlling role in resource management. According to Rodan et al. (2004), network centrality influences managers' management skills, which in turn influences project innovation performance [27]. Cappiello et al. (2020) investigated firm competitiveness in managing complicated projects using network centrality [28]. In general, network centrality is vital in coordinating and directing information transfer and knowledge sharing, and it has a significant impact on project value creation.

2.2. Social Network Analysis in Construction Project Management

Social network analysis (SNA) is an appropriate tool for examining network structure and network characteristics [4]. Existing research indicates that SNA is efficient in exploring the following areas: (1) interdependence between project-based organizations [6]; (2) interorganizational relationships [11]; (3) structural characteristics of project networks [12]; and (4) multi-level analysis of project networks [16]. Pryke (2004) emphasizes that from the perspective of stakeholders, the main focus of SNA research is to explore formal relationships between stakeholders and network structure characteristics [12]. This study focuses on the network density and network centrality of construction project networks. Network density is a fundamental indicator in SNA that quantifies the degree to which participants in a network are interconnected [18]. Mathematically, it is calculated as the ratio of the actual number of connections between participants to the maximum possible number of connections [19]. In high-density networks, participants have many direct connections, which facilitates the rapid dissemination of information, resources, and knowledge [4]. In the context of construction project networks, high-density networks can achieve seamless communication between stakeholders such as contractors, designers, and owners, potentially reducing information asymmetry and the occurrence of affective conflicts [20]. Network centrality focuses on the status and influence of individual actors in the network. It contains multiple dimensions [21]. Degree centrality calculates the number of direct connections that a participant has. A higher centrality degree indicates that participants are more embedded in the network. Betweenness centrality measures the ability of participants to control the flow of information between other participants, highlighting their role as a bridge in the network. Closeness centrality reflects the efficiency of a participant in a network reaching all other participants, typically measured by the average shortest path distance from the participant to other participants in the network [22].

Network density and network centrality are important components of SNA [20]. SNA provides a comprehensive perspective to examine the complex relationships and interactions within a construction project network. Network density provides an overview of the overall connectivity of a network, while network centrality explores the roles and influences of individual actors [9]. They help to understand how network structure shapes individual and collective behavior, which is closely related to the study of the impact of network structure on conflict and project value creation [10]. For example, high-density networks

may generate more opportunities for collaboration, but if communication channels are not effectively managed, they can also increase the likelihood of conflicts. Stakeholders with high centrality typically have significant power in coordinating project activities, mediating disputes, and driving the overall value creation process. Stakeholders with high centrality can either serve as catalysts for resolving conflicts and enhancing project value creation or escalate conflicts due to their dominant positions [20].

The relationship between building information modeling (BIM), green building project management, and SNA is profound. BIM is a revolutionary digital approach that has transformed the traditional construction industry [17]. It is a digital representation of the physical and functional aspects of buildings, creating a comprehensive 3D model-based database that integrates architecture, engineering, and construction information [19]. SNA provides a powerful toolkit to explore the complex relationships and interactive networks among stakeholders in BIM-based projects and green building projects. In the context of BIM, SNA can be used to draw information flow networks between different participants. It helps identify key nodes or influential stakeholders who play a crucial role in disseminating BIM-related knowledge and coordinating BIM-based activities [21]. In green building project management, SNA can reveal the network structure of the stakeholders involved in sustainable practice implementation. It can identify leaders who drive the adoption of green strategies, as well as weak connections or communication barriers that may hinder the project from achieving green goals. Understanding these concepts and their interrelationships is crucial for exploring the impact of network structure on conflict and project value creation. The unique network structure in BIM integration projects and green building projects can have a significant impact on the occurrence of conflicts [17]. In BIM-based projects, conflicts may arise due to issues such as data ownership, differences in BIM proficiency among team members, or unclear communication channels within the BIM network [21]. Similarly, in green building projects, conflicts may arise from different interests and priorities among stakeholders, such as developers seeking to maximize profits while environmentalists prioritize sustainability. Meanwhile, in both cases, well-structured networks can promote value creation.

Integrated project delivery (IPD) is a collaborative project delivery method in which all major stakeholders, including owners, architects, engineers, contractors, and subcontractors, participate in early integrated collaboration [10]. It has the characteristics of sharing goals, risk sharing mechanisms, and jointly pursuing value optimization throughout the entire project lifecycle [7]. SNA is a powerful analytical tool that can be effectively applied to explore complex relational networks in IPD projects. In the IPD environment, multiple stakeholders such as owners, architects, engineers, and contractors work closely together, and SNA can be used to explore the information sharing network [10]. By applying indicators such as network density and centrality, researchers can determine which stakeholders are at the core of disseminating key project knowledge and which communication channels are most effective [19]. This helps ensure the smooth flow of information within the project team and reduces the possibility of misunderstandings that may lead to conflicts. In addition, SNA can detect potential bottlenecks or weak links in collaborative networks [22]. For example, it can identify stakeholders with limited connections to others whose isolation may hinder overall cooperation and slow down project progress [21]. By recognizing these issues, appropriate measures can be taken to strengthen the network and improve project implementation efficiency.

2.3. Project Conflict

Conflict is defined as incompatibility or confrontation [29]. Project conflicts primarily appear as interorganizational conflicts in construction projects. Interorganizational con-

flicts are defined as organizational contradictions or combative actions [30]. Construction projects often entail tens of millions of dollars in investment, a lengthy construction time, a variety of technology, and complex processes [4]. Furthermore, construction projects require a huge number of project participants, such as owners, contractors, design units, consulting units, supervisory units, and financial institutions [2]. These construction organizations have varying professional skills, core capabilities, and interest demands. These elements frequently cause project conflicts. Project conflicts can arise as a result of construction schedule and resource constraints, differing understandings of project plans, and disagreements in the priority of project objectives [31]. Furthermore, one-time construction processes, incomplete contracts, the project's changing internal and external environment, and diverse cognition are all important variables that contribute to project conflicts [32].

In organizational management, conflict types often include affective conflicts involving relationships and emotions, as well as substantive conflicts involving tasks and monetary interests [20,33]. In construction projects, affective conflict reflects organizational incompatibility, manifested as negative emotions between organizations, as well as explicit conflicting behaviors such as negative confrontation, which arise when the organization feels uncoordinated or inconsistent with other organizations [34]. Substantive conflict refers to different perspectives of an organization regarding the content of a project task, as well as different perspectives, contradictions, or disagreements of the organization regarding the specific process arrangements (such as schedule arrangement, project plan formulation, and resource allocation arrangement) [35]. The impact of conflict on construction projects is multi-faceted. On the negative side, unresolved conflicts can significantly delay project schedules. Disputes over technical designs may require repeated revisions, extending the design phase and pushing back the start of construction [8]. Financially, conflicts often lead to cost overruns due to rework, legal disputes, and inefficiencies in resource allocation. In terms of quality, strained relationships among team members can result in a lack of attention to detail and corner-cutting, compromising the final product [5]. However, under certain circumstances, conflict can also have positive implications. Constructive task-related conflicts, for example, can stimulate innovation as different perspectives are debated, potentially leading to better-quality solutions.

2.4. Project Value Creation

The concept of project value creation is commonly used in fields such as economics, politics, and sociology [36]. The value creation phases in project management include the project implementation phase and the operating phase [37]. By removing non-value-added operations, a project's value can be increased. Eliminating non-value-added operations lowers construction project investment and total lifecycle costs [38]. Construction project value creation can be achieved by optimizing project implementation procedures, improving project management models, and increasing task implementation efficiency [39]. The value creation must be viewed through the eyes of several project participants (such as owners, contractors, design units, supervision units, and users), rather than just the owners or contractors [36–39].

Three procedures are used to assess the value creation of a construction project. The first step is to provide value to the project, which involves meeting the project's essential objectives (i.e., quality, duration, and cost) [40]. The second step is to add value to project stakeholders. The third step is to add value to the construction project's users, such as raising user satisfaction [41]. Based on prior research and taking into account the features of the construction industry, this study defines project value creation in two dimensions: time dimension and participant dimension. In terms of time, the project's value creation evaluation comprises both the project implementation and operation phases. Although

different project participants may have different viewpoints on project value creation, their definition of project value creation is usually consistent during the implementation and operation stages of the construction project [42]. In this study, the evaluation indicators of construction project value creation include hard element and soft element dimensions. Hard elements include quality, cost, and schedule. Soft elements include cooperative effectiveness, trust, ability enhancement, and user satisfaction [43,44].

3. Research Hypothesis and Model

3.1. Research Hypothesis

3.1.1. Density and Conflicts

Network density reflects group cohesion. High network density promotes identity and trust among network actors, facilitating engagement and information exchange [19]. Previous research has shown that network density can influence the attitudes and behaviors of network actors, consequently affecting organizational performance [20–22]. A construction project is a temporary organizational cooperation network that is formed at the start of the project and disassembled at the end [2]. Unlike long-term cooperation networks (such as those in high-tech industries), organizations in transitory cooperation networks are frequently unfamiliar with one another, restricting interorganizational interaction [4].

A high-density cooperative network is a highly interconnected communication network, which promotes communication among organizations and increases the flexibility of interorganizational interactions, achieving the coordination required during project tasks and process implementation [20]. Furthermore, a high-density cooperative network is typically more active, which aids in the exchange of interorganizational knowledge and information, the reduction of affective conflicts, and the promotion of cooperation efficiency and effectiveness [22]. A high-density project network frequently possesses a wealth of knowledge and information. Due to differences in information and resources gained by project participants, viewpoints on project task layout may differ [31]. This may lead to discrepancies in task substance or contradictions in process arrangements amongst project participants, resulting in substantive conflicts [34].

According to social network theory, a higher network density indicates a greater number of connections and more frequent interactions among project stakeholders [17]. These intensive interactions facilitate the establishment of trust-based relationships, as repeated communication and cooperation help stakeholders understand each other's values, working styles, and interests. In the realm of construction projects, when contractors, subcontractors, designers, and owners interact closely, they are more likely to resolve potential misunderstandings promptly [11]. For example, regular meetings and information exchanges can prevent minor disagreements from escalating into intense emotional confrontations. From the perspective of conflict management, positive relationships nurtured by high-density networks act as a buffer against affective conflicts [18]. When stakeholders trust and respect each other, they are more willing to adopt a collaborative attitude to address differences, rather than allowing emotions to dominate the situation [13]. In contrast, in a network with low density, limited interactions may lead to insufficient understanding, and when problems occur, stakeholders may lack the necessary emotional connection and communication channels to resolve them effectively, thus increasing the probability and intensity of affective conflicts [45].

H1a. *Substantial conflicts are positively impacted by network density.*

H1b. *Affective conflicts are negatively impacted by network density.*

3.1.2. Centrality and Conflicts

Network centrality describes the location of network actors, as well as their capacity to access and control network resources [15,23]. A highly centralized organization is positioned in the network's core and has more opportunities to connect with other organizations [25]. A highly centralized organization means a strong project-participating organization that can promote task completion through their own influence and power [26]. Meanwhile, organizations with high centrality have more links, which might foster task resource integration and organizational cooperation [24]. This frequently decreases organizational inconsistencies about project tasks and leads to the development of consistent task resolution procedures, resulting in fewer interorganizational substantive conflicts.

Social network theory posits that entities with high network centrality possess greater control over resources, information, and decision-making power within the network [12]. In the context of construction projects, stakeholders with high centrality, such as main contractors or key design firms, are in a prime position to access critical project- related information and resources [20]. This advantage enables them to identify potential sources of substantial conflicts, such as disputes over task assignments, resource allocation, or technical specifications, at an early stage. From the perspective of conflict management, these central stakeholders can leverage their influential position and abundant resources to mediate conflicts effectively [8]. For instance, a general contractor with high network centrality can organize timely coordination meetings, share comprehensive project information among subcontractors, designers, and owners, and enforce unified standards and procedures [11]. By doing so, they can facilitate communication, align the interests of different parties, and promote consensus-building, thereby reducing the likelihood and severity of substantial conflicts. Construction projects typically include three types of deliverables: DB (i.e., design–build), DBB (i.e., design–bid–build), and IPD (i.e., integrated project delivery) [2,5,8]. Owners and contractors exhibit high centrality, but this pattern is contingent on the project delivery method [20]. Specifically, owners typically dominate design-phase centrality in DBB projects, contractors lead execution-phase centrality in DB projects, and both achieve balanced high centrality in IPD projects. Organizations in the construction project network have various interest demands [20]. The owner plans the project procedure in order to shorten the construction period. The contractor must, however, examine the sequencing of each task as well as the proper allocation of funds, personnel, and machinery [29]. The high centrality of owners and contractors can have an impact on information flow and resource sharing [25]. This frequently results in debates about project process arrangements and controversies over the project's limited resource distribution, aggravating interorganizational affective conflicts. In the construction project network, although highly centralized organizations can play a role that crosses organizational boundaries, they may also create bottlenecks in information transmission and communication between organizations, because organizations with high centrality may distort information transmission and increase suspicion and distrust among other organizations [21,22]. During the construction project's implementation, this will exacerbate affective conflicts between organizations [30].

H2a. *Substantial conflicts are negatively impacted by network centrality.*

H2b. *Affective conflicts are positively impacted by network centrality.*

3.1.3. Conflict and Project Value Creation

This study divides project conflicts into two types: substantive conflicts and affective conflicts [20]. Conflicts of various types have diverse effects on project value creation.

Affective conflicts frequently have a negative impact on decision-making and interorganizational cooperation, eventually impeding project value creation [29]. There are three specific causes listed. First, affective conflicts might cause organizational members to conceal their genuine thoughts and beliefs, reducing communication and information exchange between organizations. Second, affective conflicts between organizations can lead to heated interactions and move the focus of the organization's attention away from task content, ultimately breaking down collaboration between organizations. Third, affective conflict can diminish understanding between organizations, leading to an increase in confrontation [29]. Overall, affective conflicts impede efficient communication and cooperation among organizations, thus harming project value creation.

Substantive conflicts occur during the implementation of construction projects due to discrepancies in task content or procedure configurations [34]. A great number of fresh opinions and ideas sometimes emerge when addressing project tasks, resulting in substantive conflicts. It is widely assumed that substantive conflicts aid in the attainment of project objectives [30]. This is because substantive conflicts assist organizations in recognizing and exploring diverse perspectives and opinions, hence improving their understanding of project tasks and processes. Substantive conflicts can also improve the organization's critical evaluation of task content, objectives, and execution techniques, hence driving task completion [32]. Furthermore, substantive conflicts can boost organizational cohesion, interorganizational linkages, and collaboration in the completion of difficult tasks [35]. This contributes to project value creation.

Organizational conflict theory indicates that under certain conditions, conflicts can stimulate constructive dialogue and innovation [20]. Substantive conflicts, which typically revolve around task-related issues such as technical approaches, resource allocation strategies, and project objectives, can bring diverse perspectives to the fore in construction projects [8]. For example, when different design teams hold conflicting views on the structural design of a building, one advocating for a more traditional and cost-effective approach while the other proposing an innovative but potentially more expensive solution, this conflict can trigger in-depth discussions [29]. Through such debates, stakeholders are forced to re-evaluate and justify their positions, leading to the discovery of new ideas that may integrate the advantages of both sides, such as a hybrid design that balances cost-effectiveness and innovation. Furthermore, substantive conflicts can also enhance the overall quality of project decision-making [32]. When conflicts arise, project teams are motivated to conduct more comprehensive research and analysis, involving multiple stakeholders in the decision-making process. This inclusive approach not only enriches the information base for decision-making but also increases the acceptance and commitment of stakeholders to the final decisions, reducing the likelihood of future disruptions and ensuring the smooth progress of the project towards value creation goals [20].

H3a. *Project value creation is positively impacted by substantive conflict.*

H3b. *Project value creation is negatively impacted by affective conflict.*

3.1.4. Network and Project Value Creation

A construction project is a temporary cooperative network consisting of numerous project participants [4]. An important topic in project management is the analysis of organizational output from a network perspective [20]. Construction projects are complex, uncertain, and multi-objective [8]. When making decisions, organizations must thoroughly comprehend the numerous risks, changes in the environment, and task complexity [5]. However, highly centralized organizations may restrict other organizations from expressing their thoughts and viewpoints, impacting interactions between organizations [21].

Previous research results have indicated that centrality has become a barrier to information transfer [24,35]. Organizations in the network's core may also cause information distortion, influencing other organizations' decisions and increasing their dissatisfaction. This will weaken interorganizational cohesion, reduce the efficiency and effectiveness of interorganizational cooperation, and have a negative influence on project value creation.

When compared to sparse networks, high-density project networks often have stronger cohesion, making it easier for organizations to create similar cognitive, behavioral, and normative standards [18]. These shared cognitive, behavioral, and normative criteria contribute to a more consistent basis of cooperation among organizations, facilitating project implementation and delivery [24]. Furthermore, high-density collaborative networks can reduce the opportunistic behavior of project participants to some extent [19]. This is because, in this highly interconnected cooperative network, an organization's opportunistic behavior can swiftly spread throughout the network, harming the organization's reputation. At the same time, other organizations in the network are more inclined to impose sanctions on the organization. As a result, opportunism is less likely to occur in a high-density project network [21]. It can be concluded that the advantages provided by high-density project networks are beneficial for project value creation.

H4a. *Project value creation is positively impacted by network density.*

H4b. *Project value creation is negatively impacted by network centrality.*

3.2. Research Model

The impact of project networks on conflicts and project value creation is still unknown. This study explores the impact of project networks on conflicts and project value creation using the input–mediator–output (IMO) model [20]. The input variable is network structure (i.e., network density and network centrality). From a theoretical perspective, network centrality and network density are two fundamental and widely-recognized indicators in network analysis [18,19]. Furthermore, these two indicators have been proven to be crucial in exploring the mechanisms of network-related phenomena across various fields, providing a solid theoretical basis for this study [11,17,20]. The mediating variables include two forms of conflict: affective conflict and substantive conflict. Project value creation is the output variable. Figure 1 depicts a theoretical model. The model specifically proposes that (1) network structure may lead to affective conflict and substantive conflict; (2) two types of conflict may affect project value creation; (3) network structure may affect project value creation; and (4) project conflict may act as a mediator between network structure and project value creation.

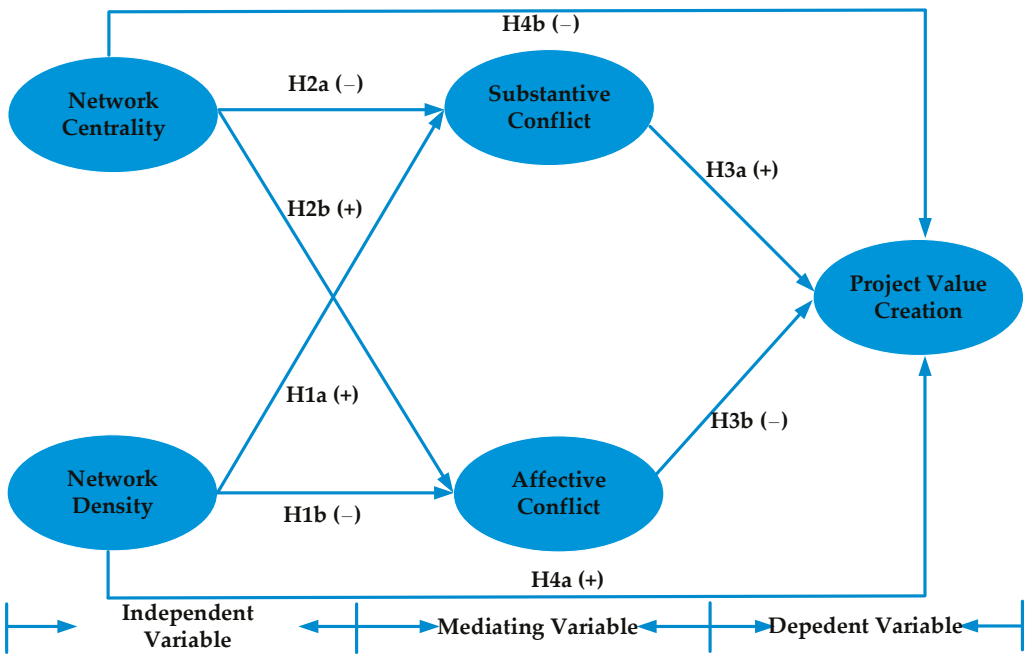


Figure 1. Research model.

4. Methods

4.1. Questionnaire

A questionnaire was created to investigate the variables. Basic demographic information was analyzed, such as work experience and job position. The measurement items include network structure items, conflict items, and project value creation items. To develop measurement items, the three stages listed below are used. The first stage is to locate and cite items. The second stage is to improve items. Specifically, the items related to network structure were adapted from the research of Friedkin (1981), who conducted in-depth studies on network analysis in construction projects [18]. His research provides a solid theoretical foundation for measuring network centrality and density. For the measurement of project conflict, we referred to the scales developed by Gardiner & Simmons (1992) and Zeng et al. (2022), which have been widely used and validated in the field of project management [8,20]. These studies focus on distinguishing different types of conflict and their impact on project performance. As for project value creation, the items were inspired by the research of Li et al. (2020), Szatmari et al. (2021), and Huang et al. (2020), who proposed a comprehensive framework for evaluating project output from multiple perspectives [13–15]. We chose these studies because of their high relevance and authority in the respective research areas, and we made appropriate modifications to ensure the items were suitable for our specific research context.

The third stage involves discussions with experts in the construction industry [15]. The items were developed based on previous research. The goal of discussions with experts is to confirm the suitability of all items [19]. We solicited the opinions of 13 experts from owners, contractors, supervisory units, and design units. All experts had over 8 years of work experience in the construction project management field. They had participated in a wide range of projects, including large-scale infrastructure construction, commercial building projects, and residential development projects. In terms of education, 6 of them held a Ph.D. degree in construction management or related fields, while the remaining 7 had a master's degree. Their educational backgrounds covered disciplines such as civil engineering, management science, and engineering economics. Regarding their positions, 5 experts were senior project managers in well-known construction enterprises, being responsible

for overall project planning, execution, and control. Four experts served as professors at leading universities, engaged in both teaching and research in construction project management. The other 4 experts were consultants at professional construction consulting firms, providing strategic advice and technical support for various construction projects. This diverse group of experts ensured a comprehensive and professional perspective in the validation of our measurement items and research model. We came to an agreement on the items' applicability after discussions. Except for the demographic information, all items in the questionnaire employed the Likert five-point scale (where 1 means "strongly disagree" and 5 means "strongly agree").

To convert these qualitative opinions into numerical values suitable for structural equation modeling (SEM) analysis, we calculated the mean and standard deviation of the expert ratings for each path. The mean value was used as the initial estimate of the path coefficient [20]. This approach is in line with the common practice in the literature when incorporating subjective judgments into SEM. We used a two-stage approach in the SEM analysis. In the first stage, we ran the model with the initial path coefficient estimates from expert opinions as starting values. Then, in the second stage, we allowed the model to freely estimate the path coefficients based on the collected data, while still respecting the theoretical relationships indicated by the expert opinions [15]. This iterative process helped to optimize the path coefficients and ensure that they reflected both the expert knowledge and the empirical evidence from the data.

4.2. Pilot Test

Pilot tests were conducted in construction projects in Shanghai, Shandong Province, and Jiangsu Province. The interviewees included professionals from owners, contractors, design units, and supervisory units. A total of 230 questionnaires were distributed. After screening 147 collected questionnaires, 96 were judged to be valid. During the questionnaire screening process, we removed questionnaires that were not answered attentively or thoroughly. We then conducted normality tests on valid samples using a quantity–quantity plot. This is a common diagnostic technique for determining whether data has a normal distribution [20]. Figure 2 shows that the sample distribution is linear, thus satisfying a normal distribution.

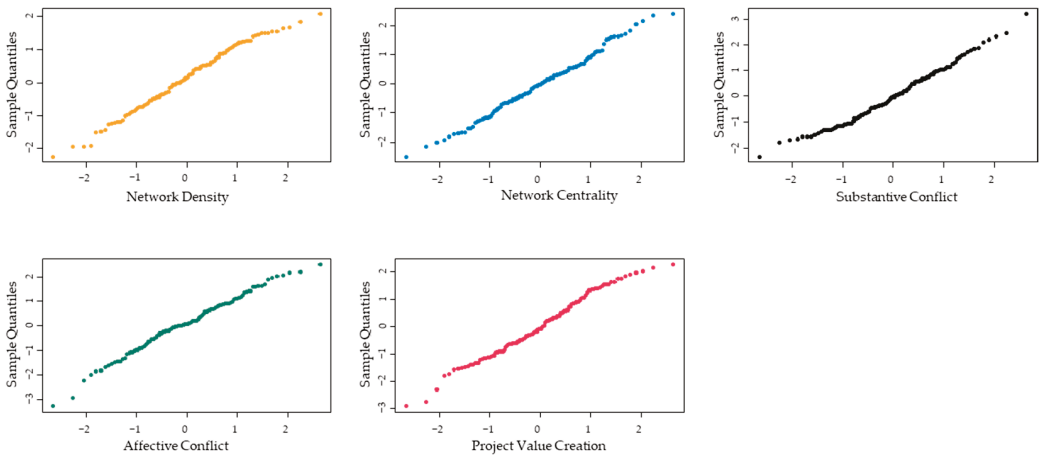


Figure 2. Sample distribution.

In the pilot test, the corrected item total correlation (CITC) coefficient and Cronbach's α coefficient were utilized to assess the reliability and validity of items [20,28]. The CITC value should not be less than 0.5. Cronbach's α coefficient should not be less than 0.7. In addition, exploratory factor analysis (EFA) was performed. The results of EFA are presented in Table 1, and the questionnaire is shown in Table 2.

Table 1. Results of EFA.

Variable	Item Code	Factor Loading	CITC Value	Cronbach's Alpha Coefficient
Network Density	ND1	0.606	0.527	0.766
	ND2	0.752	0.588	
	ND3	0.664	0.584	
	ND4	0.760	0.592	
	ND5	0.746	0.597	
Network Centrality	NC1	0.668	0.551	0.813
	NC2	0.819	0.512	
	NC3	0.727	0.602	
	NC4	0.807	0.593	
	NC5	0.652	0.676	
Substantive Conflict	SC1	0.780	0.594	0.726
	SC2	0.710	0.597	
	SC3	0.782	0.591	
	SC4	0.779	0.585	
	SC5	0.731	0.616	
Affective Conflict	AC1	0.870	0.705	0.841
	AC2	0.814	0.744	
	AC3	0.712	0.590	
	AC4	0.791	0.612	
	AC5	0.807	0.671	
	AC6	0.729	0.631	

Table 2. Questionnaire.

Variables	Measurement Items
Network Centrality	1. Other organizations within this project rely on your organization for task coordination. 2. Your organization is critical to the project's success. 3. Many organizations within this project have close relationships to yours. 4. Your organization within this project frequently serves as a go-between. 5. Among other organizations within this project, your organization has high influence.
Network Density	1. Your organization within this project has numerous direct relationships with other organizations. 2. Organizations within this project are closely tied to each other through the project network. 3. Many organizations within this project exhibit a high degree of cohesion. 4. The project network is extremely active. 5. The project network's organizations within this project have an excellent cooperation relationship.
Affective conflict	1. There is a lot of emotional tension among organizations. 2. There are several clashes among organizations. 3. There is a lot of nonconformity among organizations. 4. One party dislikes the other party's style. 5. Organizations have different cultures.
Substantive conflict	1. Organizations actively express their views. 2. Work-related issues are frequently discussed among organizations. 3. Partners present various perspectives on project task. 4. Partners frequently provide suggestions for your work. 5. Partners have differing viewpoints on process difficulties.
Project value creation	1. The quality, cost, and duration targets have been achieved. 2. The resource use efficiency is high. 3. This project has a favorable impact on users. 4. The project process has received great marks from the partners. 5. Interorganizational trust has improved. 6. Partners look forward to future cooperation.

4.3. Data Collection

Nonprobability sampling was adopted in this study. This method is widely used in construction management studies [15]. The survey samples are from various construction projects in Shanghai, Jiangsu Province, and Shandong Province, and include technical personnel, as well as middle and senior management personnel from owners, contractors, supervisory units, and design units. We distributed 800 questionnaires by email and express delivery to technical and middle and senior management professionals, and 308 questionnaires were valid, yielding a response rate of 38.5% (308/800), which is consistent with the 30–40% benchmark used in most project management research [35]. Table 3 shows the sample information. Several factors contributed to the low response rate. From an objective perspective, the construction project domain poses unique challenges for sample collection. Construction projects are often large-scale, complex, and time-consuming endeavors, with a limited number of ongoing projects at any given time. Many enterprises are also highly sensitive about disclosing project-related information due to business confidentiality and competitive concerns, which significantly restricts our access to potential respondents. Subjectively, during the survey process, there might have been issues such as overly complex questionnaire design, which could have discouraged some participants from completing the survey. Additionally, the timing of the survey might not have been optimal for some busy project managers and stakeholders, leading to a lower willingness to participate.

Table 3. Sample information.

Characteristic	Category	Frequency	%
Project category	Highway project	83	26.9
	Railway project	91	29.7
	Industrial park project	134	43.4
Work position	Project engineer	115	37.3
	Department manager	79	25.6
	Professional manager	86	27.9
	Project manager	28	9.2
Job experience	<8 years	109	35.4
	8–18 years	151	49.0
	>18 years	48	15.6
Project participating organization	Owner	81	26.3
	Contractors	98	31.8
	Design units	80	26.0
	Supervision units	49	15.9

4.4. Factor Analysis

Confirmatory factor analysis (CFA) can confirm the applicability of the items [46,47]. The CFA was calculated using AMOS 21.0. CFA created the construct reliability (CR) factors. The item that standardizes factor loads less than 0.6 was removed. The consistency between items is reflected in the CR. The CR value should be more than 0.5. The average variance extraction (AVE) method was employed to assess the convergence validity. If the AVE value is more than 0.5, the variable's item has strong convergence validity. Fit metrics such as the ratio of chi square statistics to degrees of freedom (x^2/df), the comparative fit index (CFI), and the incremental fit index (IFI) were used to test the model's quality of fit. Specifically, the x^2/df should be less than 3. The RMSEA should be less than 0.08. CFI and IFI should all be greater than 0.9 [20]. The results of CFA are presented in Table 4. While the above framework posits causal relationships between network structure, conflict,

and project value creation, empirical validation is critical to confirm these hypotheses. The following section outlines the research design to test the theoretical model using SEM.

Table 4. Results of CFA.

Variables	CR	AVE	Fit Indexes
Network density	0.81	0.71	$x^2/\mathrm{df} = 2.63$; RMSEA = 0.063; GFI = 0.92; AGFI = 0.95; NFI = 0.91; IFI = 0.92; CFI = 0.94
Network centrality	0.82	0.66	$x^2/\mathrm{df} = 1.90$; RMSEA = 0.072; GFI = 0.93; AGFI = 0.92; NFI = 0.93; IFI = 0.94; CFI = 0.91
Affective conflict	0.72	0.65	$x^2/\mathrm{df} = 1.76$; RMSEA = 0.065; GFI = 0.92; AGFI = 0.93; NFI = 0.94; IFI = 0.96; CFI = 0.95
Substantive conflict	0.73	0.63	$x^2/\mathrm{df} = 1.85$; RMSEA = 0.067; GFI = 0.92; AGFI = 0.91; NFI = 0.93; IFI = 0.91; CFI = 0.94
Project value creation	0.78	0.67	$x^2/\mathrm{df} = 1.96$; RMSEA = 0.063; GFI = 0.94; AGFI = 0.92; NFI = 0.95; IFI = 0.91; CFI = 0.93

5. Model Test of the Theoretical Model

5.1. Structural Equation Model Test

SEM is the core analytical tool for testing the hypothesized relationships between network structure, conflict, and project value creation [34]. Bootstrap, a resampling technique, was employed to confirm the mediating role of conflict [20]. Unlike traditional Sobel tests, bootstrap does not assume normality of indirect effects, making it more reliable for our context [35]. These two methods played a crucial synergistic role in the analysis of this study. Specifically, SEM estimates the direct effects between latent variables, while bootstrap method is used to estimate indirect effects.

SEM is a popular method for investigating the correlations between variables in construction project management research [48]. The theoretical model was tested using SEM in this study. The fitting indications matched the standards. The x^2/df ratio was 1.81. The RMSEA was 0.071, which is less than the criterion of 0.08. CFI and IFI are all greater than 0.9. Figure 3 and Table 5 illustrate the model test results.

Table 5. Results of SEM analysis.

Hypotheses	Coefficient	Critical Ratio	Standard Error	*p* Values
H1a	0.12 *	2.379	0.043	0.017
H1b	−0.16 ***	−4.078	0.040	0.000
H2a	−0.14 *	2.205	0.046	0.026
H2b	0.68 *	2.375	0.041	0.017
H3a	0.27 ***	5.793	0.040	0.000
H3b	−0.31 ***	−5.640	0.051	0.000
H4b	−0.12 **	−2.596	0.059	0.008
H4a	0.18 *	2.307	0.047	0.031

Notes: Abbreviations: ND, network density; SC, substantive conflict; AC, affective conflict; NC, net- work centrality; PVC, project value creation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The SEM test results are shown in Table 5. First, network density positively impacts substantive conflict and project value creation, providing support for H1a and H4a, respectively (ND → SC, 0.12; ND → PVC, 0.18). Second, the impact of network centrality on substantive conflict and project value creation is negative, providing support for H2a and H4b, respectively (NC → SC, −0.14; NC → PVC, −0.12). Third, the impact of density on affective conflict is negative (ND → AC, −0.16), while the impact of centrality on affective

conflict is positive (NC → AC, 0.68). This provides support for H1b and H2b, respectively. Fourth, the impact of substantive conflict and affective conflict on project value creation is positive and negative, respectively (SC → PVC, 0.27; AC → PVC, −0.31), which supports H3a and H3b. Furthermore, the direct, indirect, and total effects of indicators on project value creation are shown in Table 6.

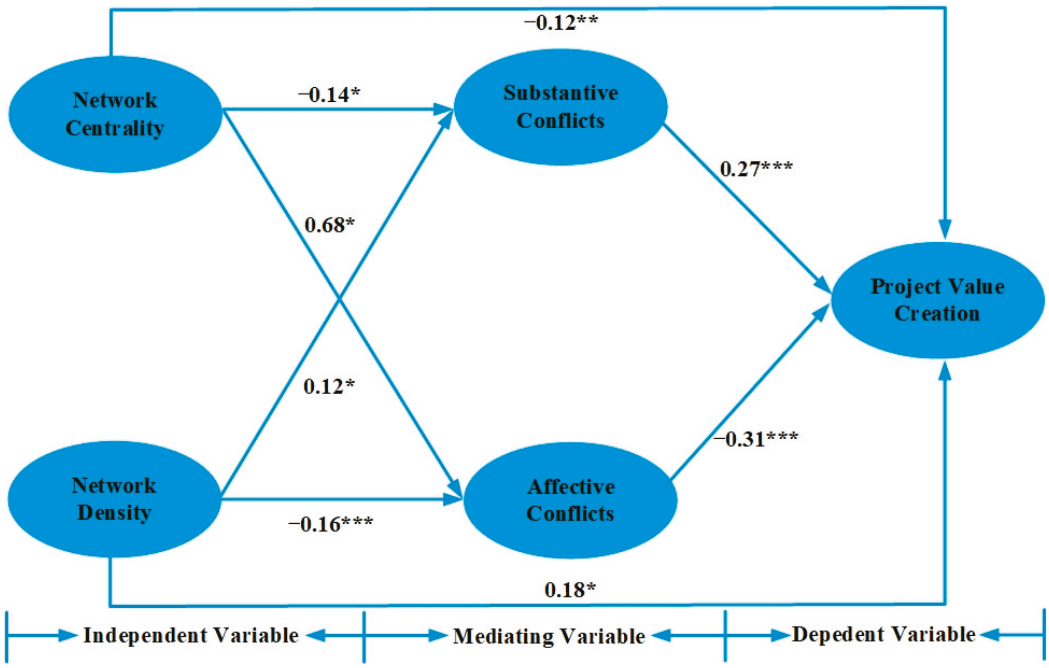


Figure 3. Model results. Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 6. Direct, indirect, and total effects of indicators on project value creation.

Indicator	Direct Effect	*p*-Value	Indirect Effect	*p*-Value	Total Effect	*p*-Value
Network Density	0.18 *	0.03	0.32 **	0.01	0.50 ***	0.000
Network Centrality	−0.12 **	0.008	−0.25 *	0.03	−0.37 **	0.001
Substantive Conflict	0.27 ***	0.000	—	—	0.27 ***	0.000
Affective Conflict	−0.31 ***	0.000	—	—	−0.31 ***	0.000

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.2. Sub-Sample Analysis

This study stratified the sample into three groups based on stakeholder count (measured as the number of organizations involved): small projects (<10 stakeholders, $n = 28$); medium projects (10–20 stakeholders, $n = 35$); large projects (>20 stakeholders, $n = 17$). This study ran separate SEMs for each group. The results show that the path coefficients of each group are significant (see Table 7).

Table 7. Sub-sample SEM results.

Path	Small Projects (n = 28)	Medium Projects (n = 35)	Large Projects (n = 17)	Overall Sample (n = 80)
Network Density → Affective Conflict	−0.12 **	−0.21 **	−0.25 *	−0.16 **
Network Centrality → Substantive Conflict	−0.22 *	−0.26 **	−0.18 *	−0.14 *
Affective Conflict → Project Value Creation	−0.35 **	−0.30 *	−0.28 *	−0.31 ***
Substantive Conflict → Project Value Creation	0.18 *	0.22 *	0.25 **	0.27 **
Network Density → Project Value Creation	0.22 **	0.11 *	0.16 *	0.18 *
Network Centrality → Project Value Creation	−0.17 *	−0.14 **	−0.18 *	−0.12 **

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Stakeholder count groups: small = <10 stakeholders; medium = 10–20 stakeholders; large = >20 stakeholders.

5.3. Measurement Invariance Testing

This study conducted multi-group confirmatory factor analysis (MGCFA) to ensure measurement equivalence across project groups. The results show configural, metric, and scalar invariance (CFI > 0.90, RMSEA < 0.08), confirming that "network structure", "conflict", and "project value creation" are measured consistently across small, medium, and large projects (see Table 8).

Table 8. Multi-group confirmatory factor analysis (MGCFA).

Invariance Type	CFI (Network Structure)	RMSEA (Network Structure)	CFI (Interorganizational Conflict)	RMSEA (Interorganizational Conflict)	CFI (Project Value Creation)	RMSEA (Project Value Creation)
Configural Invariance	0.92	0.04	0.93	0.06	0.94	0.05
Metric Invariance	0.91	0.04	0.90	0.03	0.92	0.02
Scalar Invariance	0.93	0.03	0.91	0.03	0.92	0.04

5.4. Mediation Test

After the SEM analysis, this study conducted a mediation effect analysis [49,50]. The bootstrap sample size was set to 3000 [51]. The results (Table 9) show that (1) substantive conflict mediates the relationship between network structure and project value creation (0.125, CI = [0.160, 0.271]; −0.272, CI = [0.115, 0.292]); and (2) affective conflict mediates the relationship between network structure and project value creation (−0.430, CI = [0.102, 0.241]; 0.319, CI = [0.131, 0.304]). According to the above results, project conflict serves as a mediator.

Table 9. Results of mediation test.

Source	Coefficients		CI	
	Estimate	Boot Standard Error	Lower	Upper
Substantive conflict				
Between network density and project value creation	0.125	0.026	0.160	0.271
Between network centrality and project value creation	−0.272	0.021	0.115	0.292
Affective conflict				
Between network density and project value creation	−0.430	0.019	0.102	0.241
Between network centrality and project value creation	0.319	0.034	0.131	0.304

6. Discussions

6.1. Network and Conflicts

According to the research results, network centrality promotes affective conflicts while decreasing substantive conflicts. On the contrary, network density increases substantive conflicts while decreasing affective conflicts. It can be seen that network structure has a diverse impact on conflicts. The finding that network centrality exacerbates affective conflicts is consistent with the research conclusion of Zhao and Jiang (2021), which suggests that high-centrality networks not only lead to power struggles, but also to information distortion [52]. This creates circumstances for the escalation of affective conflicts. High-centrality organizations have influence and authority in the construction project, not only affecting interactions between organizations but also managing information transfer, leading to distrust among different organizations throughout project implementation [53, 54]. This results in affective conflicts throughout project implementation. The result that network centrality reduces substantive conflicts is similar to the research conclusion of Wang et al. (2015), implying that high-centrality networks encourage consistent activity among project participating organizations [55].

The findings of this study show that network density increases substantive conflicts while decreasing affective conflicts. This is because information and expertise are disseminated quickly and efficiently in high-density project networks, which promotes interorganizational interaction and the resolution of affective conflicts [56]. A densely connected network promotes information sharing inside the network and the establishment of common normative standards between organizations, hence boosting cooperation among project participants [57]. Furthermore, tightly connected project networks can improve network flexibility and reduce tension between organizations. This is due to the fact that densely connected networks can improve interorganizational interaction, communication, and trust, as well as promote the diversity of resources and information in the network [58]. A highly connected project network also aids in the generation of distinct network clusters. Different knowledge bases, professional competencies, and varying interest demands among network clusters might contribute to substantive conflicts.

To provide a deeper interpretation of the research findings, this study employed a qualitative interview design. Firstly, this study identified 11 interview participants, including project managers, contractors, designers, and subcontractors, as their different roles and perspectives in the project network can provide comprehensive insights [20]. Secondly, semi-structured interview questions were developed. These questions were open-ended and aimed to explore various dimensions related to the research question. For example, we asked participants to describe how they perceive the network density and centrality within their projects. We inquired about specific instances of substantial and affective conflicts they had experienced, especially in situations where network density was high or certain stakeholders had significant centrality. The questions also covered how the network structure influenced the emergence, development, and resolution of these conflicts. Additionally, we asked about their understanding of the mechanisms through which network density and centrality might impact different types of conflict.

Based on our preliminary analysis of the interview data, we can already propose some initial findings regarding the internal logic. It appears that high network density may suppress affective conflict primarily by enhancing trust between organizations. Frequent interactions in a dense network provide more opportunities for stakeholders to demonstrate reliability, fulfill commitments, and understand each other's values and working styles, thereby fostering trust [13]. This increased trust, in turn, serves as a buffer against emotional tensions, reducing the likelihood of affective conflicts [31]. Based on our preliminary analysis of the interview data, several key findings emerged. When it

came to network centrality, participants described situations where stakeholders with high centrality often had significant control over resources and decision-making. This power concentration sometimes led to dissatisfaction and resistance from other parties, triggering affective conflicts, such as disputes over power and influence. In terms of substantial conflicts, the dominance of centrally-positioned stakeholders could result in less diverse perspectives being considered, potentially leading to suboptimal decisions and a decrease in task-related conflicts. In conclusion, qualitative interviews provided us with valuable insights into the complex relationships between network structure, substantial conflicts, and affective conflicts.

6.2. The Impact of Project Conflicts

The findings of this study show that substantive conflicts have a positive impact on project value creation, while affective conflicts have a negative impact. These results are consistent with prior studies, which revealed that substantive conflict, as a constructive conflict, improves organizational performance [59,60]. In addition, this study's findings highlight project conflict as a mediator between network structure and project output. Affective conflict has a stronger mediating effect than substantive conflict.

Affective conflicts commonly impede interorganizational cooperation and project task completion, limiting project value creation [61]. Furthermore, affective conflicts can shift an organization's focus away from project tasks and toward interorganizational relationships. This can have an impact on organizational cognitive function, interorganizational cooperation, and task implementation, as well as cause interorganizational conflict and impede project task implementation, resulting in a negative impact on project value creation [62]. Substantive conflicts increase the generation of new ideas, which speeds up the settlement of project problems [63]. Substantive conflicts can also lead to the evaluation of multiple disciplinary areas, as well as in-depth thinking and creative ideas on project tasks. This can promote interorganizational cooperation, and thereby project value creation.

To provide a deeper interpretation of the research findings, this study employed a qualitative interview approach. Firstly, this study identified 11 interview participants, including project managers, contractors, designers, and subcontractors. Each of these roles plays a distinct part in project value creation and conflict management, and their varied perspectives provided a comprehensive understanding of the research topic [20]. Semi-structured interview questions were developed. These questions are open-ended and aim to explore various dimensions related to the research question. For example, we asked participants to describe specific instances of substantive conflict they had encountered, such as disputes over technical specifications or resource allocation, and how these conflicts were resolved. We inquired about the immediate and long-term impacts of such conflicts on the project, including whether they led to innovations, improved decision-making, or enhanced the overall quality of the project, thereby contributing to value creation. Regarding affective conflict, we explored how emotional tensions, personality clashes, or power struggles within the project team affected communication, collaboration, and ultimately, project performance and value. The questions covered stakeholders' definitions and evaluations of project value creation and how they perceive the role of conflicts in this process.

Based on our preliminary analysis of the interview data, we found that substantive conflict often acts as a driver for project value creation. Stakeholders frequently reported that intense discussions and disagreements over technical or operational issues led to the discovery of more efficient construction methods, better-optimized resource allocation plans, and innovative design solutions [2]. These outcomes directly contributed to cost savings, improved project quality, and enhanced market competitiveness, thereby increasing project value. On the contrary, affective conflict was consistently described as a hindrance.

The participants noted that emotional disputes damaged team morale, caused communication breakdowns, and led to delays in decision-making and task execution [11]. Such negative consequences ultimately eroded project value by increasing costs, compromising project schedules, and reducing stakeholder satisfaction [30]. In conclusion, the qualitative interviews enabled us to uncover the nuanced relationships between different types of conflict and project value creation.

6.3. Network and Project Value Creation

The research conclusion indicates that network centrality negatively impacts project value creation, lending credence to the notion that high network centrality is detrimental to organizational output. This result contradicts the findings of Li et al. (2020), who discovered that network centrality positively impacts project performance in interorganizational cooperation networks with high-tech backgrounds [64]. This is due to the fact that the high-tech industry is distinct from the construction industry. Cooperation between organizations in the high-tech industry is long-term, whereas collaboration in a construction project is short-term [20]. In a construction project, the collaborative connection starts at the outset and ends when the project is completed. Building emotional trust amongst organizations is challenging, since construction projects are transient and one-time in nature [20]. In this case, the stature and power that high-centrality organizations possess have the potential to sow mistrust among other organizations, endangering interorganizational collaboration [22]. Furthermore, information transmission can be distorted by high-centrality organizations, which can have a negative impact on interorganizational communication and information transmission [9]. These elements have a detrimental effect on project tasks, which in turn negatively impact project value creation.

The findings of this study show that network density positively impacts project value creation. This research conclusion is similar to previous research findings [65]. This is because construction projects are a temporary network composed of different project-participating organizations. A tightly connected construction project network makes it easier for organizations to share information and creates a unified action standard, which encourages collaboration across project-participating organizations [66]. Networks with strong connections can lessen tensions between organizations, boost organizational flexibility, and encourage interorganizational cooperation [67]. Thus, a tightly connected project network facilitates the efficient execution of construction projects and project value creation.

7. Theoretical Implications

This research has the following theoretical implications. First, the core theoretical implication of this study lies in the integrated framework that bridges network structure, conflict types, and project value creation—an under-explored linkage in the construction management literature. Unlike existing studies that either focus on network structure alone or conflict management in isolation, this study theoretically proposes and empirically validates the mediating role of conflict types (substantive conflict and affective conflict) in the network structure–value relationship. This is particularly novel in the high-conflict construction context, where prior research has overlooked how network density/centrality differentially trigger conflict subtypes, thereby affecting project value creation.

Second, prior studies investigating construction project networks have predominantly relied on bivariate regression or simple correlation analysis to explore direct relationships (e.g., between network density and project performance). These methods are ill-equipped to unpack mediating processes—a critical gap, as theoretical frameworks in project management increasingly emphasize that network structure influences outcomes through intermediate factors (e.g., conflict) rather than directly. In contrast, this study employs

SEM, a multivariate technique uniquely suited for testing complex mediation models with multiple latent variables. By specifying a theoretical framework where network structure (density/centrality) predicts conflict types (substantive conflict/affective conflict), which in turn predict project value creation, this study formally validates the chain of effects hypothesized in our model. SEM allows us to estimate both direct effects and indirect effects (e.g., density → conflict → project value creation) simultaneously, using bootstrapping to confirm the significance of mediation paths. This approach addresses a key limitation of prior work, which often assumed but failed to statistically validate such mediating roles.

Third, this study demonstrates the detrimental effects of network centrality on project value creation, as well as the positive influence of network density on project value creation, by presenting evidence on the relationship between network structure and project value creation. These findings broaden our understanding of the construction project network. This study's findings also highlighted conflict's positive and destructive consequences. Furthermore, this study investigated the mediating role of conflicts in the setting of construction project networks. The findings revealed that substantive conflicts would mitigate the detrimental impact of project network structure on project value creation, while affective conflicts would increase it. As a result, a construction project network with a high degree of substantive conflict and a low level of affective conflict can give conditions for organizations to span divergent factions to boost project value creation.

8. Practical Implications

This study has the following practical significance. For policy: this study has practical significance for policy-making in the construction industry. Based on the results of this study, the current policies and regulations in the construction industry can be analyzed, and targeted improvement suggestions can be proposed. For example, it is suggested to establish industry standards for the design of network structures in construction projects. Developing specific guidelines for network density and centrality can help project participants avoid potential risks caused by inappropriate network structures. In addition, it is recommended that incentive policies be introduced to encourage construction companies to adopt effective conflict management strategies. For example, providing financial subsidies or preferential treatment to companies implementing comprehensive conflict resolution mechanisms in project bidding can promote the overall improvement of industry management practices. Integrating the results of this research into policy-making can help create a more scientific and regulatory friendly environment for the construction industry.

For society, construction projects play a crucial role in social development, and this study contributes to bringing positive social impacts. Optimizing the network structure of construction projects and effectively managing conflicts within them can help promote social harmony. By more effectively allocating resources within a well-organized project network, social resource waste can be reduced. By emphasizing these social benefits, this study aims to attract more social attention and support, and promote sustainable social development. In addition, project managers must understand that various types of project disputes have varying degrees of impact on project value creation. To promote the healthy development of the construction project network, the constructive impact of substantive conflicts should also be considered, and appropriate measures should be taken, such as brainstorming meetings that encourage different organizations to participate in discussions on project task implementation and process arrangements, thereby deepening understanding of task content and implementation methods and promoting smooth project implementation. In addition, project managers should be aware of the destructive effects of affective conflicts and take necessary actions, such as creating a pleasant project culture and learning environment, to mitigate their negative impact on society.

9. Conclusions

This study investigates the relationship between network structure, project conflicts, and project value creation in the construction project network. The results show that (1) network centrality is not conducive to resolving affective conflicts or promoting project value creation; (2) network density promotes the settlement of affective conflicts and project value creation; (3) substantive conflicts have a beneficial impact on project value creation, whereas affective conflicts have a negative impact; and (4) project conflict mediates the process of network structure influencing project value creation. This study serves as a reference for network structure governance, conflict governance, and project value management in construction projects. This study contributes to our existing understanding of the effect of project network structure on project value creation in the field of construction project management. This provides a new direction for construction project governance and project value management. Furthermore, this study validated the constructive role of substantive conflicts and the destructive impact of affective conflicts, contributing to the conflict governance literature.

10. Limitations

This study has the following limitations. First, different types of conflict might morph into each other under particular conditions. The transformation process between conflicts was not investigated in this study. Future research should investigate the metamorphosis of the two forms of conflict and their impact on project value creation. Second, the execution of construction projects is a dynamic process, and the project network structure may alter over time. As a result, future study should investigate the dynamic changes in project network topology and their impact on project value creation. Third, the sample data for this study is limited to Chinese construction projects. Future research should include more construction projects in order to broaden the research findings.

Author Contributions: Conceptualization, J.C.; Methodology, C.L. and Y.S.; Software, C.L. and J.C.; Validation, Y.S. and J.C.; Formal Analysis, C.L.; Investigation, Y.S.; Data Curation, C.L.; Writing—original draft preparation, C.L.; Writing—review and editing, Y.S.; Visualization, C.L.; Supervision, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kasprowicz, T.; Starczyk-Kołbyk, A.; Wójcik, R.R. The randomized method of estimating the net present value of construction projects efficiency. *Int. J. Constr. Manag.* **2023**, *23*, 2126–2133. [CrossRef]
2. Cheraghi, E.; Zohrehvandi, S.; Vanhoucke, M.; Tosarkani, B.M. A multiproject scheduling and resource management model in projects construction. *Eng. Constr. Archit. Manag.* **2023**, *30*, 1578–1600. [CrossRef]
3. Wuni, I.Y.; Shen, G.Q. Exploring the critical success determinants for supply chain management in modular integrated construction projects. *Smart Sustain. Built Environ.* **2023**, *12*, 258–276. [CrossRef]
4. Zheng, X.; Le, Y.; Chan, A.P.C.; Hu, Y.; Li, Y. Review of the application of social network analysis (SNA) in construction project management research. *Int. J. Proj. Manag.* **2016**, *34*, 1214–1225. [CrossRef]
5. Ökmen, Ö.; Öztaş, A. Construction project network evaluation with correlated schedule risk analysis model. *J. Constr. Eng. Manag.* **2008**, *134*, 49–63. [CrossRef]
6. Shabani Ardakani, S.; Nik-Bakht, M. Social network analysis of project procurement in Iranian construction mega projects. *Asian J. Civ. Eng.* **2021**, *22*, 809–829. [CrossRef]

7. Maskil-Leitan, R.; Reychav, I. A sustainable sociocultural combination of building information modeling with integrated project delivery in a social network perspective. *Clean Technol. Environ. Policy* **2018**, *20*, 1017–1032. [CrossRef]
8. Gardiner, P.D.; Simmons, J.E.L. Analysis of conflict and change in construction projects. *Constr. Manag. Econ.* **1992**, *10*, 459–478. [CrossRef]
9. Xue, X.; Zhang, R.; Wang, L.; Fan, H.; Yang, R.J.; Dai, J. Collaborative innovation in construction project: A social network perspective. *KSCE J. Civ. Eng.* **2018**, *22*, 417–427. [CrossRef]
10. Hickethier, G.; Tommelein, I.D.; Lostuvali, B. Social network analysis of information flow in an IPD-project design organization. In Proceedings of the International Group for Lean Construction, Fortaleza, CE, Brazil, 29 July–2 August 2013.
11. Pryke, S.D. Towards a social network theory of project governance. *Constr. Manag. Econ.* **2005**, *23*, 927–939. [CrossRef]
12. Pryke, S.D. Analysing construction project coalitions: Exploring the application of social network analysis. *Constr. Manag. Econ.* **2004**, *22*, 787–797. [CrossRef]
13. Li, J.; Jiang, W.; Zuo, J. The effects of trust network among project participants on project performance based on SNA approach: A case study in China. *Int. J. Constr. Manag.* **2020**, *20*, 837–847. [CrossRef]
14. Szatmari, B.; Deichmann, D.; van den Ende, J.; King, B.G. Great successes and great failures: The impact of project leader status on project performance and performance extremeness. *J. Manag. Stud.* **2021**, *58*, 1267–1293. [CrossRef]
15. Huang, Y.; Shi, Q.; Pena-Mora, F.; Lu, Y.; Shen, C. Exploring the impact of information and communication technology on team social capital and construction project performance. *J. Manag. Eng.* **2020**, *36*, 04020056. [CrossRef]
16. Pryke, S.; Badi, S.; Almadhoob, H.; Soundararaj, B.; Addyman, S. Self-organizing networks in complex infrastructure projects. *Proj. Manag. J.* **2018**, *49*, 18–41. [CrossRef]
17. Maskil-Leitan, R.; Gurevich, U.; Reychav, I. BIM management measure for an effective green building project. *Buildings* **2020**, *10*, 147. [CrossRef]
18. Friedkin, N.E. The development of structure in random networks: An analysis of the effects of increasing network density on five measures of structure. *Soc. Netw.* **1981**, *3*, 41–52. [CrossRef]
19. El-Diraby, T.; Krijnen, T.; Papagelis, M. BIM-based collaborative design and socio-technical analytics of green buildings. *Autom. Constr.* **2017**, *82*, 59–74. [CrossRef]
20. Zeng, H.; Cao, J.; Fu, Q. Unpacking the "Black Box": Understanding the effect of strength of ties on inter-team conflict and project success in megaprojects. *Buildings* **2022**, *12*, 1906. [CrossRef]
21. Cao, D.; Li, H.; Wang, G.; Luo, X.; Yang, X.; Tan, D. Dynamics of project-based collaborative networks for BIM implementation: Analysis based on stochastic actor-oriented models. *J. Manag. Eng.* **2017**, *33*, 04016055. [CrossRef]
22. Tang, Y.; Wang, G.; Li, H.; Cao, D. Dynamics of collaborative networks between contractors and subcontractors in the construction industry: Evidence from national quality award projects in China. *J. Constr. Eng. Manag.* **2018**, *144*, 05018009. [CrossRef]
23. Ibarra, H. Network centrality, power, and innovation involvement: Determinants of technical and administrative roles. *Acad. Manag. J.* **1993**, *36*, 471–501. [CrossRef]
24. El-Khatib, R.; Fogel, K.; Jandik, T. CEO network centrality and merger performance. *J. Financ. Econ.* **2015**, *116*, 349–382. [CrossRef]
25. Lee, S.H.M.; Cotte, J.; Noseworthy, T.J. The role of network centrality in the flow of consumer influence. *J. Consum. Psychol.* **2010**, *20*, 66–77. [CrossRef]
26. Mullen, B.; Johnson, C.; Salas, E. Effects of communication network structure: Components of positional centrality. *Soc. Netw.* **1991**, *13*, 169–185. [CrossRef]
27. Rodan, S.; Galunic, C. More than network structure: How knowledge heterogeneity influences managerial performance and innovativeness. *Strateg. Manag. J.* **2004**, *25*, 541–562. [CrossRef]
28. Cappiello, G.; Giordani, F.; Visentin, M. Social capital and its effect on networked firm innovation and competitiveness. *Ind. Mark. Manag.* **2020**, *89*, 422–430. [CrossRef]
29. Jaffar, N.; Tharim, A.H.A.; Shuib, M.N. Factors of conflict in construction industry: A literature review. *Procedia Eng.* **2011**, *20*, 193–202. [CrossRef]
30. Bruce, P.J.; Hrymak, V.; Bruce, C.; Byrne, J. The role of interpersonal conflict as a cause of work-related stress in construction managers in Ireland. *Constr. Innov.* **2025**, *25*, 61–77. [CrossRef]
31. Fenn, P.; Lowe, D.; Speck, C. Conflict and dispute in construction. *Constr. Manag. Econ.* **1997**, *15*, 513–518. [CrossRef]
32. Zhang, L.; Huo, X. The impact of interpersonal conflict on construction project performance: A moderated mediation study from China. *Int. J. Confl. Manag.* **2015**, *26*, 479–498. [CrossRef]
33. Li, T.H.Y.; Ng, S.T.; Skitmore, M. Conflict or consensus: An investigation of stakeholder concerns during the participation process of major infrastructure and construction projects in Hong Kong. *Habitat Int.* **2012**, *36*, 333–342. [CrossRef]
34. Chen, Y.Q.; Zhang, Y.B.; Zhang, S.J. Impacts of different types of owner-contractor conflict on cost performance in construction projects. *J. Constr. Eng. Manag.* **2014**, *140*, 04014017. [CrossRef]
35. Tang, Y.; Chen, Y.; Hua, Y.; Fu, Y. Impacts of risk allocation on conflict negotiation costs in construction projects: Does managerial control matter? *Int. J. Proj. Manag.* **2020**, *38*, 188–199. [CrossRef]

36. Laursen, M.; Svejvig, P. Taking stock of project value creation: A structured literature review with future directions for research and practice. *Int. J. Proj. Manag.* **2016**, *34*, 736–747. [CrossRef]
37. Willumsen, P.; Oehmen, J.; Stingl, V.; Geraldi, J. Value creation through project risk management. *Int. J. Proj. Manag.* **2019**, *37*, 731–749. [CrossRef]
38. Hjelmbrekke, H.; Klakegg, O.J.; Lohne, J. Governing value creation in construction project: A new model. *Int. J. Manag. Proj. Bus.* **2017**, *10*, 60–83. [CrossRef]
39. Winter, M.; Szczepanek, T. Projects and programmes as value creation processes: A new perspective and some practical implications. *Int. J. Proj. Manag.* **2008**, *26*, 95–103. [CrossRef]
40. Haddadi, A.; Johansen, A.; Andersen, B. A conceptual framework to enhance value creation in construction projects. *Procedia Comput. Sci.* **2016**, *100*, 565–573. [CrossRef]
41. Borgonovo, E.; Gatti, S.; Peccati, L. What drives value creation in investment projects? An application of sensitivity analysis to project finance transactions. *Eur. J. Oper. Res.* **2010**, *205*, 227–236. [CrossRef]
42. Artto, K.; Ahola, T.; Vartiainen, V. From the front end of projects to the back end of operations: Managing projects for value creation throughout the system lifecycle. *Int. J. Proj. Manag.* **2016**, *34*, 258–270. [CrossRef]
43. Lee-Kelley, L.; Sankey, T. Global virtual teams for value creation and project success: A case study. *Int. J. Proj. Manag.* **2008**, *26*, 51–62. [CrossRef]
44. Hsu, J.S.C.; Chang, J.Y.T.; Klein, G.; Jiang, J.J. Exploring the impact of team mental models on information utilization and project performance in system development. *Int. J. Proj. Manag.* **2011**, *29*, 1–12. [CrossRef]
45. Sparrowe, R.T.; Liden, R.C.; Wayne, S.J.; Kraimer, M.L. Social networks and the performance of individuals and groups. *Acad. Manag. J.* **2001**, *44*, 316–325. [CrossRef]
46. Borkenau, P.; Ostendorf, F. Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personal. Individ. Differ.* **1990**, *11*, 515–524. [CrossRef]
47. Anderson, J.C.; Gerbing, D.W. Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *J. Appl. Psychol.* **1991**, *76*, 732. [CrossRef]
48. Yoon, Y.; Gursoy, D.; Chen, J.S. Validating a tourism development theory with structural equation modeling. *Tour. Manag.* **2001**, *22*, 363–372. [CrossRef]
49. Haq, M.A.; Abbasi, S. Indirect impact of hedonic consumption and emotions on impulse purchase behavior: A double mediation model. *J. Manag. Sci.* **2016**, *3*, 108–122. [CrossRef]
50. Hayes, A.F.; Montoya, A.K.; Rockwood, N.J. The analysis of mechanisms and their contingencies: PROCESS versus structural equation modeling. *Australas. Mark. J.* **2017**, *25*, 76–81. [CrossRef]
51. Tahir, M.S.; Ahmed, A.D.; Richards, D.W. Financial literacy and financial well-being of Australian consumers: A moderated mediation model of impulsivity and financial capability. *Int. J. Bank Mark.* **2021**, *39*, 1377–1394. [CrossRef]
52. Zhao, H.; Jiang, J. Role stress, emotional exhaustion, and knowledge hiding: The joint moderating effects of network centrality and structural holes. *Curr. Psychol.* **2021**, *41*, 1–13. [CrossRef]
53. Hossain, L. Effect of organisational position and network centrality on project coordination. *Int. J. Proj. Manag.* **2009**, *27*, 680–689. [CrossRef]
54. Ibarra, H.; Andrews, S.B. Power, social influence, and sense making: Effects of network centrality and proximity on employee perceptions. *Adm. Sci. Q.* **1993**, *38*, 277–303. [CrossRef]
55. Wang, H.; Zhao, J.; Li, Y.; Li, C. Network centrality, organizational innovation, and performance: A meta-analysis. *Can. J. Adm. Sci.* **2015**, *32*, 146–159. [CrossRef]
56. Yuan, J.; Chen, K.; Li, W.; Ji, C. Social network analysis for social risks of construction projects in high-density urban areas in China. *J. Clean. Prod.* **2018**, *198*, 940–961. [CrossRef]
57. Xu, L.; Ding, R.; Wang, L. How to facilitate knowledge diffusion in collaborative innovation projects by adjusting network density and project roles. *Scientometrics* **2022**, *127*, 1353–1379. [CrossRef]
58. Zou, X.; Ingram, P.; Higgins, E.T. Social networks and life satisfaction: The interplay of network density and regulatory focus. *Motiv. Emot.* **2015**, *39*, 693–713. [CrossRef] [PubMed]
59. Euwema, M.C.; Van de Vliert, E.; Bakker, A.B. Substantive and relational effectiveness of organizational conflict behavior. *Int. J. Confl. Manag.* **2003**, *14*, 119–139. [CrossRef]
60. Afzalur Rahim, M. Toward a theory of managing organizational conflict. *Int. J. Confl. Manag.* **2002**, *13*, 206–235. [CrossRef]
61. Vahtera, P.; Buckley, P.; Aliyev, M. Affective conflict and identification of knowledge sources in MNE teams. *Int. Bus. Rev.* **2017**, *26*, 881–895. [CrossRef]
62. Parayitam, S.; Dooley, R.S. The interplay between cognitive-and affective conflict and cognition-and affect-based trust in influencing decision outcomes. *J. Bus. Res.* **2009**, *62*, 789–796. [CrossRef]
63. Elbanna, S. The impact of affective conflict on firm performance. *Manag. Res. News* **2009**, *32*, 789–803. [CrossRef]

64. Li, M.; Xiao, F.; Cheng, Y.; Xie, B.; Liu, C.; Xu, B. Exploring the relationship between network position and innovation performance: Evidence from a social network analysis of high and new tech companies from a less-developed area in China. *Chin. Manag. Stud.* **2020**, *14*, 93–111. [CrossRef]
65. Tian, M.; Su, Y.; Yang, Z. Exploration versus exploitation: The influence of network density on firm's strategic choice between two types of innovation. *Technol. Anal. Strateg. Manag.* **2022**, *36*, 1–14. [CrossRef]
66. Walker, M.H. The contingent value of embeddedness: Self-affirming social environments, network density, and well-being. *Soc. Ment. Health* **2015**, *5*, 128–144. [CrossRef]
67. Luarn, P.; Chiu, Y.P. Influence of network density on information diffusion on social network sites: The mediating effects of transmitter activity. *Inf. Dev.* **2016**, *32*, 389–397. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

How Does Port Logistics Service Innovation Enhance Cross-Border e-Commerce Enterprise Performance? An Empirical Study in Ningbo-Zhoushan Port, China

Weitao Jiang [1], Hongxu Lu [1], Zexin Wang [2] and Ying Jing [1,*]

[1] Business School, NingboTech University, Ningbo 315100, China; jiangwt2007@nit.zju.edu.cn (W.J.); zjulhx@zju.edu.cn (H.L.)
[2] College of Economics, Zhejiang University, Hangzhou 310000, China; wangzexin@zju.edu.cn
* Correspondence: y.crystal@nit.zju.edu.cn; Tel.: +86-15067436297

Abstract: The port logistics service innovation (PLSI) is closely associated with cross-border e-commerce (CBEC) enterprise performance, given that the port, as the spatial carrier and the joint point of goods, information, customs house affairs, etc., is essentially a key node of the CBEC logistics chain. However, the influence mechanism of PLSI on CBEC enterprise performance has still not yet been elaborated by consensus. To fill this gap, this study aims to figure out the effect mechanism integrating the probe into two variables (i.e., information interaction and environmental upgrade) in a moderated mediation model. Specifically, this study collects questionnaire survey data of logistics enterprises and CBEC enterprises in the Ningbo-Zhoushan Port of China by the Bootstrap method in the software SPSS 26.0. The results show the following: (1) PLSI can positively affect the CBEC enterprise performance; (2) information interaction plays an intermediary role between PLSI and CBEC enterprise performance; and (3) environmental upgrade can not only positively regulate the relationship between information interaction and CBEC enterprise performance, but also enhance the mediating role of information interaction with a moderated intermediary effect.

Keywords: port logistics service innovation (PLSI); cross-border e-commerce (CBEC) enterprise performance; information interaction; environmental upgrade; moderated mediating model

1. Introduction

The combination of global consumers' online shopping habits and China's strong commodity supply capacity has made cross-border e-commerce (CBEC) a new model for Chinese enterprises to "enter overseas markets" [1]. Ports are at the hub node in the logistics chain of CBEC; port logistics service enterprises that rely on ports to provide logistics services have become the key to promote the efficient operation of cross-border transportation at this node. Port logistics service enterprises are "specialized service organizations in the port economic circle". They create value through symbiosis and evolution with port operators, shipping companies, CBEC, and other entities [2]. Refined cross-border market competition and personalized CBEC enterprise demand are increasingly challenging the era of reform and innovation of port logistics services. How the innovation of port logistics services affects the "black box" of CBEC business performance has gradually become the focus of scholars [3–5].

The existing research on port logistics service mainly focuses on two aspects: The first is to pay attention to the innovation mechanism of port logistics service. The research on the innovation mode of port logistics service enterprises shows a paradigm shift from linear innovation to networked collaboration. The main innovation mechanisms include the following: (1) Technology-driven innovation. Focusing on single-point breakthroughs in technology, such as building a path optimization model, it has been proven that algorithm-driven transportation route adjustment can increase the punctuality rate by 22% and reduce the cost by 15%. Such studies ignore the organizational adaptability [6]; (2) Institutional synergy innovation. Enterprises' participation in the "single window" policy design can significantly improve the service efficiency, reduce the customs clearance time by 50%, and reduce compliance costs by 33%, revealing the leverage of government enterprise system co-creation on service innovation [7]; (3) Ecological-enabling innovation. The relevant research turned to the construction of innovation ecology. The 4PL platform integrated the resources of enterprises, governments, and R&D institutions through the triple-helix model, which increased the service iteration speed of small- and medium-sized logistics enterprises by 2.1 times and reduced the customer churn rate by 18% [8]. The second is to focus on the interaction mechanism between port logistics services and service objects. The research on the interaction of port logistics services presents a paradigm shift from one-way supply to ecological co-creation. Most scholars discuss how the quality of interaction between port owners and operators, as key subjects connecting the global supply chain and regional economy, and service objects (freight forwarders, cross-border e-commerce, manufacturers, etc.) affect the efficiency of international trade from the macro level [9–11]. The empirical research shows that the logistics timeliness is improved by 0.63 units for each unit of resource complementarity between port and cargo owner enterprises, which verifies the core proposition of "integrating Operand and Operant resources to create value" [12]; it also puts forward that the core feature of the third-generation port is to shift from "cargo handling" to "whole process value-added services", and the service object participates in value creation through information sharing, process collaboration, etc. [13].

There are three disjunctions in the existing research, which restrict the depth of theoretical development. The first is the dislocation of the research object: 78% of the literature equated macro-level port operators with micro-level "port logistics service enterprises" (third-party enterprises that rely on ports to provide logistics services, such as freight forwarders, warehousing service providers, multimodal transport operators, etc., but not port operators and not port owners), confused port facility owners and port service providers, and lacked a deep description of the interaction between micro enterprises, resulting in the obscuration of the innovation mechanism of third-party port logistics service enterprises. The second is the weak theoretical foundation. There is a lack of academic explanation such as a value co-creation theory for "collaborative behavior" such as cost sharing and revenue sharing between port logistics enterprises and CBEC enterprises. The third is the lack of internalization of environmental factors. Most works of literature focus on static resource integration; emergency innovation in turbulent environments has not been theorized; and there is a lack of in-depth discussion on the "uncertainty of business environment" in cross-border logistics.

From a practical perspective, geographical conflicts such as the Red Sea crisis have led to a surge in the risk of traditional shipping routes. The interruption rate of international shipping routes has increased by 40%, and the inventory cost of CBEC has surged by 25% [14]. The real crisis of cross-border logistics forces theoretical innovation, further highlighting the urgency of revealing "the complex superposition effect of key exogenous process variables such as information interaction and environmental upgrade on the direct linkage mechanism between port logistics service enterprises and CBEC enterprises".

The Ningbo-Zhoushan port (Figure 1) is located in Ningbo and Zhoushan, Zhejiang Province, China (Latitude: 29.92959° N, Longitude: 121.844481° E), and comprises 19 port areas including Beilun, Yangshan, Liuheng, Qushan, and Chuanshan, etc. It is a first-class port opened to the outside world. For 15 consecutive years, its cargo throughput ranks first in the world, and its container throughput ranks among the first three in the world. It is also the fourth 30 million class port in the world after the Shanghai port, Singapore port, and Shenzhen port. Around the Ningbo-Zhoushan port, many well-known port logistics service enterprises and CBEC clusters have gathered, forming a well-established industrial chain and supply chain system. Therefore, it is of typical significance to explore how the service innovation of port logistics service enterprises can activate the business performance of CBEC based on the Ningbo-Zhoushan port.

Figure 1. Ningbo-Zhoushan port: (**a**) left: northwestward; (**b**) middle; and (**c**) right: southeastward.

This study takes port logistics service enterprises and CBEC enterprises in the Ningbo-Zhoushan port cross-border logistics chain as the research object, based on the service-oriented value co-creation theory, and introduces "information interaction" as the intermediary variable and "environmental upgrade" as the moderator variable, so as to build a moderated intermediary model, and explore the impact mechanism of port logistics service innovation (PLSI) on the performance of CBEC enterprises. The possible contributions of this study are as follows: the first is to expand the research boundary of value co-creation theory, and reveal that port logistics service enterprises (non-port-operators and non-port-owners) create value with CBEC enterprises in a dynamic environment through the chain mechanism of "resource reorganization → information integration → environment adaptation → value output", providing a theoretical paradigm for third-party port logistics service enterprises to break through resource constraints and strengthen core competitiveness. The second is to promote the development of a service-dominant logic process theory, and reveal that information resources reconstruct the path of value co-creation through "information interaction", and service innovation needs to penetrate the organizational boundary through "information interaction", modify the static cognition of "innovation

directly drives performance", and build a new theoretical base of dynamic competition and cooperation. The third is to deepen the innovation of service-oriented value co-creation theory, and reveal that value co-creation has contextual variability; "environmental upgrade" adjusts the efficiency of value co-creation by changing the original boundary conditions of enterprise operation and reconstructing the rules of resource integration, and promotes the refined development of a service-oriented value co-creation theory in the dimension of "information interaction–environmental upgrade". The research framework is shown in Figure 2.

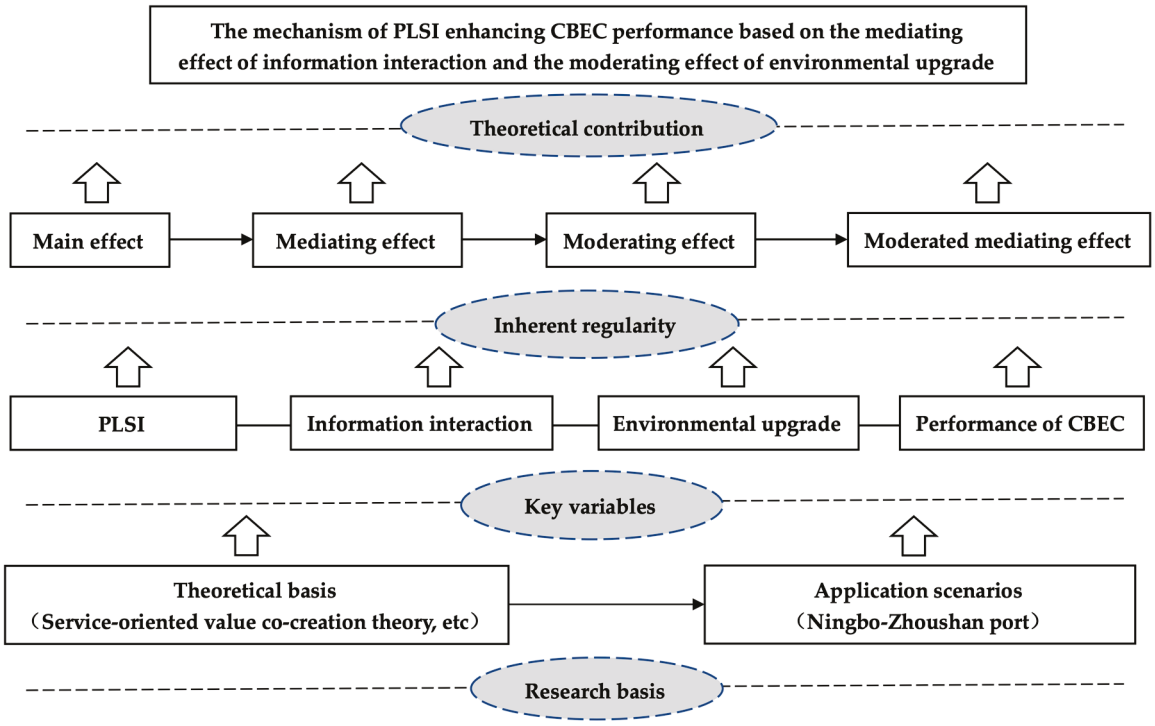


Figure 2. Research framework.

2. Theoretical Framework and Research Hypothesis

2.1. Theoretical Framework

The service-oriented value co-creation theory is the integration of the service-dominant logic and value co-creation theory, emphasizing that "economic value comes from service exchange". The essence of co-creation is that value is not provided unilaterally by enterprises, but is defined and realized jointly by service providers and users in the interaction [15]. Its core is to integrate Operand and Operant resources to create value (Table 1); that is, the real innovation value does not lie in the number of port facilities (Operand), but in how to transform these facilities into service solutions to customer pain points through data, algorithms, and collaboration mechanisms (Operant).

Table 1. Resource classification.

Resource Type	Essential Attribute	Performance in Port Logistics
Operand resources	Operated static entity resources	Port infrastructure (wharf, crane), goods, vehicles, etc.
Operant resources	Dynamic capability resources that can actively create value	Logistics algorithm, employee skills, customer data, collaboration process, etc.

Specifically, for port logistics service enterprises that rely on the Ningbo-Zhoushan port to carry out business, the essence of service innovation is to dynamically integrate two types

of core resources: first, Operant resources (dynamic resources), including the enterprise's digital capability (e.g., order forecasting algorithm), customer collaboration mechanism (e.g., sharing inventory data with cross-border e-commerce), and process optimization experience (e.g., LCL inspection scheme); and, second, Operand resources (object resources), including leased port facilities (e.g., Ningbo Meishan bonded warehouse), cross-border means of transport (e.g., the carriage of the China Europe Express), and customs resources (such as the right to use the places under customs supervision). When port logistics service enterprises reorganize Operant resources through service innovation, the potential value of Operand resources can be activated. Compared with the traditional commodity led logic, the theoretical applicability of this theory in the port logistics service scenario is shown in Table 2.

Table 2. Adaptability of service-oriented value co-creation theory.

Dimension	Traditional Commodity Dominant Logic	Service-Oriented Value Co-Creation Logic
Subject of value creation	Unilateral creation by producers	Multi-party collaboration and co-creation
Resource role	Operand resources as the core	Operant resource drives Operand resource
Source of value creation	Value arising from displacement of goods (physical utility)	Value comes from solving the pain points of CBEC supply chain (Such as timeliness and cost)
Competitive focus	Port Facility Scale	Resource integration and responsiveness
Port logistics service cases	Port logistics service enterprises only provide standardized logistics services (loading and unloading services, etc.)	Port logistics service enterprises and CBEC enterprises jointly design service schemes (Such as "overseas warehouse + front warehouse" ecology, etc.)

Grounded in the theoretical framework of service-oriented value co-creation, value is co-created through "dynamic capabilities acting on physical resources", and its value creation mechanism is as follows: port logistics service enterprises contribute Operant resources (such as intelligent scheduling algorithms), CBEC enterprises (service objects) contribute Operant resources (such as market demand data, collaborative feedback, etc.), and then both parties jointly activate Operand resources (such as port logistics facilities, goods, etc.), thus generating new value (such as logistics timeliness improvement, cost reduction, etc.), as shown in Figure 3.

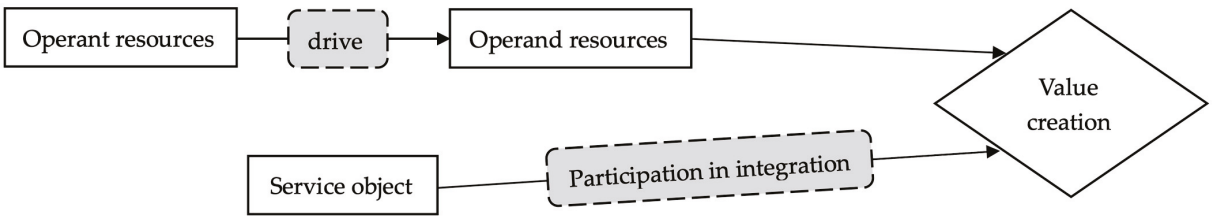


Figure 3. Service-oriented value co-creation mechanism.

This section logically demonstrates the underlying mechanism through which PLSI enhances CBEC enterprises performance. The analysis centers on core variables—PLSI, information interaction, environmental upgrade, and CBEC performance—examining their multilayered relationships through main effects, mediating effects, and moderating effects.

2.2. Research Hypothesis

2.2.1. The Mutual Effect Between PLSI, Information Interaction, and CBEC Enterprise Performance

In terms of the interaction between PLSI and the performance of CBEC enterprises, firstly, the theoretical essence of service innovation is the strategic reorganization of Operant

resources. The service-oriented value co-creation theory believes that economic value originates from the dynamic integration of Operant resources, which generates new value by acting on Operand resources. The essence of port logistics service innovation is the strategic reorganization of its Operant resources (such as digital capabilities, collaborative algorithms, process design knowledge, etc.) by port logistics service enterprises [16]. This reorganization activates the potential effectiveness of Operand resources (such as port facilities, transportation tools, etc.) through the reconstruction of the enterprise's internal capabilities. Secondly, the value co-creation mechanism is embodied in resource integration. The service-oriented value co-creation theory requires that value creation must be realized through the resource integration of service providers and users [17]. The service innovation (Operant resource reorganization) of port logistics service enterprises provides a new resource integration interface for CBEC. CBEC injects its own resources (such as commodity flow and capital flow) into the interface, and is coupled with the Operant resources of port logistics service enterprises to form a collaborative resource bundle. The resource bundle generates indivisible collaborative value (such as shortening the logistics cycle, reducing the risk of cargo damage, etc.) through mutual adaptation. Thirdly, collaborative value is transformed into performance. The service-oriented value co-creation theory proposes that performance is the perception and realization of co-created value by beneficiaries [18]. As direct beneficiaries, CBEC enterprises have effectively improved their business performance under the causal chain of "Operant restructuring → resource integration → collaborative value → performance perception", as shown in Figure 4.

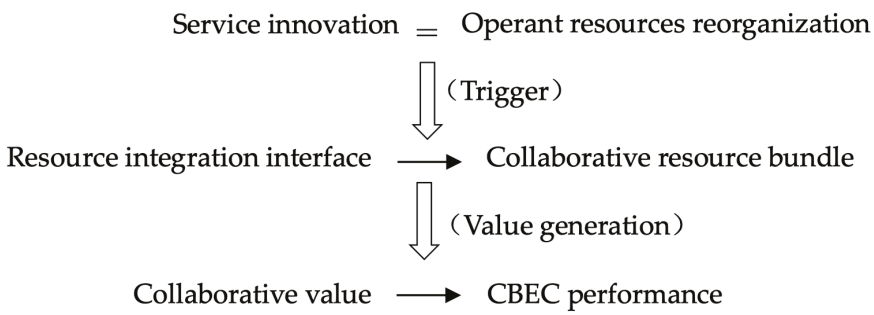


Figure 4. The logic of PLSI to improve the performance of CBEC enterprises.

To sum up, PLSI builds a resource integration interface through Operant resources reorganization, and generates collaborative value in the original resource coupling with CBEC, which is perceived by CBEC in the form of performance. Therefore, this study puts forward the following research hypothesis:

H1a: *PLSI has a significant positive impact on the performance of CBEC enterprises.*

In terms of the interaction between PLSI and information interaction, firstly, the essence of service innovation is Operant resources reorganization. The service-oriented value co-creation theory puts forward that the essence of all economic exchanges is service exchange, and services are driven by Operant resources (knowledge, skills, systems, etc.). In essence, PLSI is the strategic reorganization of its Operant resources (such as data algorithm, process design ability, collaborative knowledge, etc.) by port logistics service enterprises. The reorganization process expands the resource integration boundary of the enterprise [19] and creates a new interface for information interaction. Secondly, the reorganization of Operant resources has spawned new demands for information interaction. "Resource complementarity principle" believes that high-order Operant resources (such as an intelligent scheduling algorithm) need to be coupled with complementary resources to

release value [20]. The service innovation of port logistics service enterprises (such as a dynamic position system) requires real-time order flow data (Operant resources of CBEC) to calibrate the model, while CBEC needs to obtain visual information of the shipping schedule (Operant resources of port logistics service enterprises) to optimize inventory decisions. Therefore, the efficiency release of Operant resources after service innovation depends on "two-way information interaction"; otherwise, the resource integration interface will fail. Thirdly, service innovation reconstructs the situational rules of information interaction. The context dependence principle of the service-oriented value co-creation theory emphasizes that value co-creation is defined by a specific context [21]. A context change forces participants to adjust the information interaction mode, while PLSI reconstructs the context through three paths: (1) the complexity is increased, new variables are introduced into service innovation, and finer grained data needs to be exchanged; (2) the timeliness is compressed, and service innovation accelerates the process, forcing the frequency of information interaction to upgrade; and (3) risk redistribution and service innovation change the responsibility boundary and require information transparency improvement, as shown in Figure 5.

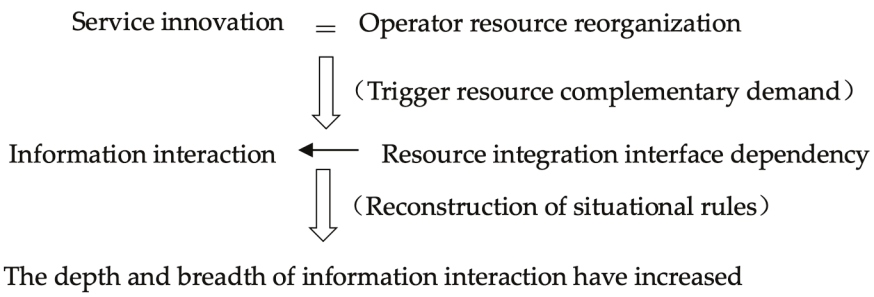


Figure 5. The logic of PLSI to information interaction.

To sum up, the PLSI expands the resource integration boundary and reconstructs the situational rules by reorganizing the Operant resources, thus endogeneously requiring the synchronous upgrading of information interaction in granularity, frequency, and depth. Thus, the following research hypothesis is formulated:

H1b: *PLSI has a significant positive impact on information interaction.*

In terms of the interaction between information interaction and the performance of CBEC enterprises, firstly, the essence of information interaction is the flow of Operant resources. The service-oriented value co-creation theory emphasizes that all economic exchanges are fundamentally service exchanges, and services transmit value through Operant resources (knowledge, skills, information, etc.). Information exchange is not simply data transmission, but the two-way flow of Operant resources: port logistics service enterprises output process knowledge (such as changes in customs clearance rules, etc.) and dynamic capabilities (such as class visualization algorithm, etc.); and CBEC enterprises input demand signals (such as promotion cycle) and feedback knowledge (such as return strategy). This process realizes the complementary integration of high-level Operant resources [22]. Secondly, resource integration has given birth to inseparable collaborative value. The inevitability principle of value co-creation holds that value is created jointly by service providers and users and realized through resource integration. Information interaction generates collaborative value through three paths: (1) the efficiency value, such as the real-time matching of shipping date data and order flow, reducing the timing friction of cross-border logistics links (such as the decline of port detention time); (2) the decision value, such as the change information of customs clearance policy, which

drives CBEC enterprises to adjust procurement strategies and reduce bullwhip effect; and (3) the risk-hedging value, real-time sharing of abnormal events (such as weather delays), and activation of collaborative emergency response between the two sides (such as the relocation of the China Europe train). Thirdly, we have the transformation mechanism from collaborative value to performance. The value situational principle of service-oriented value co-creation theory holds that value is defined and realized by beneficiaries in specific situations. As a direct beneficiary, the business performance of CBEC enterprises has been improved in efficiency, cost, risk control, and many other aspects, as shown in Figure 6.

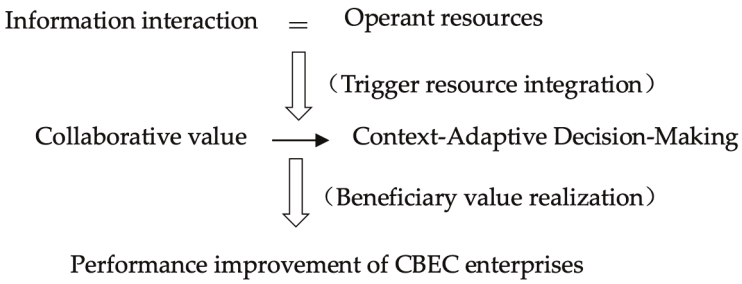


Figure 6. The logic of information interaction on business performance of CBEC enterprises.

To sum up, information interaction generates indivisible collaborative value by driving the integration of Operant resources, which is transformed into efficiency, cost, risk control, and other performance improvements by CBEC enterprises in the context of decision-making. Therefore, this study puts forward the following research hypothesis:

H1c: *Information interaction has a significant positive impact on the performance of CBEC enterprises*.

2.2.2. The Mediating Effect of Information Interaction

As an important factor of production and operation, information resources can be obtained from the transmission and sharing of information to improve the level of service innovation, improve financial income, and formulate plans. Information exchange refers to the information exchange mode in which the owner of information resources communicates with stakeholders and jointly manages specific information resources for a certain purpose and within a certain period of time [23]. Value co-creation requires both sides to continuously exchange key resources. The essence of information interaction is the process of exchange of Operant resources between port logistics service enterprises and cross-border e-commerce enterprises: port logistics service enterprises provide shipping schedule dynamics, customs clearance status, position visibility, etc., and CBEC enterprises feed back order urgency, return strategy, promotion plan, etc. Only when the Operant resources (such as data, algorithms, etc.) of both sides are deeply integrated can PLSI be transformed into CBEC business performance.

According to the theory of limited resources, enterprises cannot obtain long-term benefits only by relying on their own specific resources. Active information exchange and sharing with other market participants can help enterprises create inimitable scarce resources in combination with their own characteristics. Especially in the market with rapidly changing external environment, information and knowledge from customers provide new space for enterprise innovation [24], but, at the same time, we should pay attention to the matching of innovative services and customer needs. Through the information exchange and communication in the early stage of service innovation, we can obtain in a timely manner customer information and perceive the market dynamics, so as to activate the first mover advantage of enterprises and occupy the market share first [25]. On the one hand, high-quality information interaction with CBEC enterprises can stimulate the willingness of port logistics enterprises to innovate services and improve the possibility of success.

Smooth information exchange with CBEC enterprises provides port logistics enterprises with timely and accurate customer needs, enabling port logistics enterprises to formulate accurate customized services according to the market changes and consumption trends of CBEC, so as to maintain stable customer stickiness and market share. On the other hand, high-frequency information exchange and sharing with port logistics enterprises can enable CBEC enterprises to participate in the development process of new logistics services in real time, and accelerate the speed of PLSI [26]. The efficient logistics brought by service innovation will substantially improve the performance of CBEC enterprises. Therefore, with the help of the bridge of "information interaction" between port logistics enterprises and CBEC enterprises, PLSI has effectively improved the business performance of CBEC enterprises. Therefore, this study puts forward the following research hypothesis:

H2: *Information interaction plays an intermediary role between PLSI and CBEC enterprise performance*.

2.2.3. The Moderating Effect of Environmental Upgrade

Environmental upgrade is manifested as the external conditions affecting the production and operation of enterprises which have been in a dynamic process of unstable change degree and uncertain change scope, including the overall change in economic environment, the change of competitors' business models, the emergence of new technology application scenarios, and the upgrade or downgrade of consumer preferences. The dynamic evolution of the environment promotes enterprises' dynamic demand for information exchange [27,28]. Whether CBEC enterprises can strengthen the information interaction with port logistics enterprises, customers, and other stakeholders in the context of continuous upgrades of the environment will become the key to improving their business performance. When the environment changes relatively smoothly, the existing consumer trends, market structure, competitors, technology, equipment, and other external conditions change little, and the quantity and quality of the existing information sharing between CBEC and stakeholders can maintain the demand for cross-border e-commerce production and operation, and will not have a significant impact on business performance. With the continuous intensification of environmental changes, it means that the external conditions of the above impact on the e-commerce business have undergone quantitative or even qualitative changes, so it is urgent to grasp the latest information to adjust the production and operation mode and cope with the new situation. At this time, the information interaction between cross-border e-commerce and stakeholders becomes more urgent and important. By expanding the boundaries of information sharing and improving the quality of information sharing, cross-border e-commerce can obtain key information resources needed to re-establish the core competitive advantages in new scenarios. Therefore, environmental upgrade plays a regulating role in the impact of information interaction on the performance of CBEC enterprises. This study proposes the following hypothesis:

H3: *Environment upgrade positively moderates the relationship between information interaction and CBEC enterprise performance*.

In addition, based on the above assumptions, this study established a moderated mediation model; that is, environment upgrade moderates the mediating effect of PLSI on the CBEC enterprise performance through information interaction; that is, there is a moderated intermediary. Aprisma et al. (2020) explored the multiple relationships between environmental changes and firm performance and found that, in the influencing process of information interaction on firm performance, the final effect was significantly affected by external conditions such as market structure fluctuations and consumption

trend changes [29]. Changes in the external environment will change the original boundary conditions of the enterprise operation, directly trigger enterprises to search for new business model parameter information, and further promote the information interaction between enterprises and the upstream and downstream of the industrial chain [30,31]. Specifically, when the environmental upgrade is weak, port logistics enterprises lack the motivation for innovation due to the relatively stable business environment. The existing logistics services can meet the needs of cross-border e-commerce, so they will not frequently exchange information with CBEC enterprises. At this time, information sharing between the two sides will not have a significant impact on the performance of CBEC enterprises. On the contrary, when the environmental upgrade is more intense, port logistics enterprises must develop new service models to respond to environmental changes, resulting in more and more innovation risks for port logistics enterprises during the development cycle. This will promote port logistics enterprises to strengthen the real-time information exchange with CBEC enterprises and understand in a timely manner customer needs, to hedge the uncertainty in the innovation process of port logistics services. In addition, high environment upgrades will accelerate the emergence and diffusion of new information, which not only provides pioneering ideas and opportunities for port logistics enterprises to accelerate the upgrading and updating of existing service models but also promotes the information interaction between port logistics enterprises and CBEC enterprises. The full communication between the two sides on new service ideas and processes will greatly improve the pertinence of port logistics innovation to cross-border e-commerce, thus effectively improving the CBEC business performance [32]. To sum up, the stronger the environmental upgrade, the greater the pressure of logistics service innovation, resulting in the port logistics enterprises and CBEC enterprises in the same cross-border logistics chain further strengthening information communication, improving the efficiency of information sharing, and, finally, achieving the business objectives of rapidly responding to customer needs to expand the market share and increase financial returns. Therefore, the following research hypothesis is proposed:

H4: *Environmental upgrade positively moderates the mediating role of information interaction in the relationship between PLSI and CBEC enterprise performance.*

Therefore, based on the mediating effect of information interaction and the moderating effect of environmental upgrade, this study constructs a moderated mediation model of the impact of PLSI on the performance of CBEC enterprises, as shown in Figure 7:

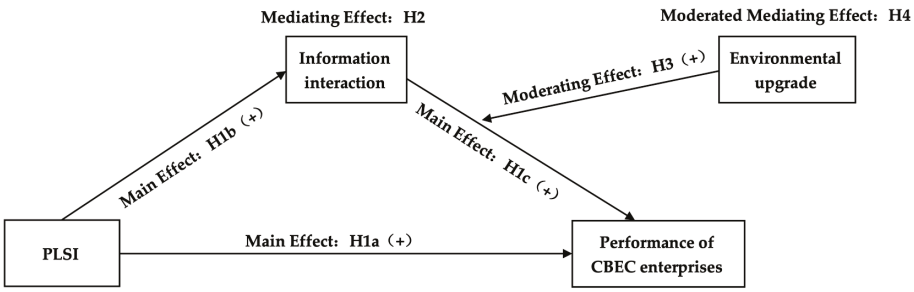


Figure 7. Theoretical model.

3. Research Design

3.1. Sampling Method

This study adopts the method of stratified random sampling combined with systematic sampling. The specific implementation process is as follows:

The first step is to establish a sampling framework. Firstly, based on the type of enterprises, the sample enterprises to be surveyed in Ningbo-Zhoushan port are divided into four independent sub-populations (strata): warehousing logistics enterprises, transportation logistics enterprises, comprehensive logistics enterprises, and CBEC enterprises. Secondly, establish an employee sampling frame, and obtain a list of all employee numbers from the human resources department of each enterprise as the basic sampling frame to ensure that all functional departments (operations/customs/sales, etc.) are covered.

The second step is to perform random sampling. The systematic sampling method is adopted within each enterprise: firstly, calculate the sampling interval k (if the total number of employees in an enterprise is 500 and the planned sampling is 25, then k = 500 ÷ 25 = 20). Secondly, determine the random starting point, and use the random number generator to select the starting number between 1 and k (for example, the starting point is 7 when k = 20). Finally, take samples at equal intervals, and take 1 employee every k people from the starting point (e.g., sample No. = 7, 27, 47, 67, . . .).

The third step is quality control measures. The criteria for determining the invalid questionnaire were as follows: (1) the completeness of the questionnaire was <90%; (2) the options of 10 consecutive questions were the same; and (3) the logic was contradictory.

3.2. Questionnaire Design

The questionnaire design adopts a three-stage process: pilot testing → refinement → finalization.

Phase I: Pilot testing. Firstly, design the initial questionnaire. Based on the service-oriented value co-creation theory, this paper constructs the core potential variables—PLSI, information interaction, environmental upgrade, and CBEC enterprise performance—designs 20 items, and uses the Likert 5-point scale. Secondly, conduct small-sample pilot testing. In the informal survey, 12 enterprises (6 port logistics service enterprises + 6 CBEC enterprises) were selected, and 36 employees were selected according to the proportion of departments: (1) cognitive interview asks the subjects' understanding of the core words one by one; (2) it takes 18.5 min on average (target $\leq$ 15 min) to complete the test, and 5 redundant items are deleted; and (3) open feedback was collected so that the dimension of "port logistics service innovation" lacks the index of "innovation mode and concept", and added 2 questions.

Phase II: Refinement. The sample size was expanded to 200 questionnaires from 16 enterprises (182 were effectively recovered), and the statistical test was as follows (Table 3):

Table 3. Reliability and validity test.

Inspection Index	Standard	Preliminary Test Results	Corrective Measures
Cronbach's α	>0.7	Environmental upgrade scale α = 0.68	Deleted 2 items with factor loadings < 0.5
KMO	>0.7	KMO of PLSI = 0.67	Consolidated 3 semantically redundant items
Factor loading	>0.6	Four items exhibited factor loadings < 0.55	Revised semantically ambiguous items
AVE	>0.5	AVE = 0.48 for information interaction	Added measurement items

Phase III: Finalization. On the basis of strictly following the feedback in the pilot testing and refinement, the finalization process realizes the optimization of measurement tools through dual integration: firstly, the questionnaire framework is systematically reconstructed to strengthen the content validity; and, secondly, multi-dimensional deviation prevention mechanism is implanted to ensure data quality. Through optimization, this study finally formed a questionnaire with both theoretical rigor and practical applicability.

3.3. Data Sources

The data of this study are derived from a questionnaire survey conducted with 62 employees of Ningbo-Zhoushan port (including 28 logistics enterprises and 34 CBEC enterprises) in 2022. A total of 400 questionnaires were issued, 346 were recovered, and the recovery rate of the questionnaires was 86.5%. After excluding invalid questionnaires such as incomplete filling and completely consistent answers, 310 valid questionnaires were obtained, with an effective rate of 89.6%. Preliminary statistical analysis of the data was performed using SPSS 26.0, and the sample distribution characteristics are shown in Tables 4 and 5.

Table 4. Sample enterprises characteristics.

Statistical Variables	Class	Frequency	Percentage (%)
Years of Operation	Less than 2 years	9	14.5
	2–5 years	28	45.2
	5–10 years	21	33.9
	More than 10 years	4	6.4
Enterprises' Size	Less than 50 people	5	8.1
	In the range of 50–100 persons	29	46.8
	Of 101–200 persons	24	38.7
	More than 200 people	4	6.4
Enterprises' Types	Warehousing-type logistics enterprises	5	8.1
	Transportation and logistics enterprises	7	11.25
	Comprehensive logistics enterprise	16	25.85
	CBEC enterprises	34	54.8
Amount To		62	100

Table 5. Sample employee characteristics.

Statistical Variables	Class	Frequency	Percentage (%)
Sex	Man	232	74.8
	Woman	78	25.2
Age	Under 30	83	26.8
	30–39 years old	163	52.6
	Age 40–49	43	13.9
	Aged 50 and over	21	6.7
Record of Formal Schooling	Junior college	110	35.6
	Undergraduate course	166	53.4
	Master	30	9.8
	Other	4	1.2
Position	Grassroots staff	216	69.7
	Middle managers	79	25.5
	Senior managers	15	4.8
Amount To		62	100

Regarding enterprises' operation years, startups (<2 years) constituted a minority (14.5%). The majority of enterprises had 2–5 years (45.2%) or 5–10 years (33.9%) of operation, while established enterprises (≥10 years) accounted for 6.4%. Regarding enterprises size, enterprises with fewer than 50 employees constituted 8.1% of the sample. Those with 50–100 employees formed the largest group (46.8%), followed by enterprises with 101–200 employees (38.7%). Enterprises exceeding 200 employees accounted for 6.4%. Regarding enterprises types, the sample included 28 port logistics enterprises (45.2%), with further breakdown: comprehensive logistics (25.85%), warehousing-based logistics (8.1%), and transportation-focused logistics (11.25%). CBEC enterprises constituted the remaining 54.8% (n = 34).

A total of 310 employees filled out the questionnaire, 232 males, 74.8%, and 78 women, 25.2%. In terms of age, 26.8% were under 30, and most of the respondents were

30–39 (52.6%), 13.9% were 40–49, and 6.7% were 50 and above. In terms of education, junior college accounted for 35.6%, with the largest number of undergraduates, accounting for 53.4%, Master's degree for 9.8%, and others for 1.2%. In terms of positions, there are many grassroots employees, accounting for 69.7%, middle managers for 25.5%, and senior managers for 4.8%. The above data show that most of the research objects in this study are young and middle-aged grassroots employees and middle managers with higher education in port logistics enterprises and CBEC enterprises. They are active in thinking and are in the front line of business, and are more able to accept and perceive innovative things, which is in line with the research theme of this study.

3.4. Variable Measurement

There were four research variables in this study, including PLSI, information interaction, environmental upgrade, and the performance of CBEC enterprises. The measurement scales of variables are based on the mature scale of the existing literature; combined with the discussion and suggestions of experts and scholars and the recent research using the scale, and, after pilot testing and refinement, they finally form a questionnaire on the impact of PLSI on the performance of CBEC enterprises. Details are as follows:

Learn from the 4-dimensional service innovation model proposed by Pim et al. (2001) (i.e., new technology selection, new service concept, new delivery system, and new customer interface) [33]; combined with Calantone et al. (2002) proposed logistics service innovation scale [34] and applications of the scale in recent studies (Jing et al., 2020) [35], this study adopts the idea that PLSI is based on the new logistics service concept, with new logistics technology as the means, the service process, and management mode, and makes logistics service value appear in a new form to meet the diversified needs of customers. Therefore, PLSI is measured, including 4 items (Table 6).

Taking the conclusions of Ennew (1999) [36] and Yen et al. (2004) [37], and the measurement index developed by McEvily (2005) [38] and applications of the scale in recent studies (Bin et al., 2020) [39] into consideration, by screening the measurement dimensions of the indicators, and taking into account the current situation of information interaction between logistics service enterprises and CBEC enterprises in Ningbo-Zhoushan port, this study puts forward that information interaction is the Internet technology and customer information exchange behavior, including market dynamic sharing, business process communication, innovation, value experience feedback, etc. Therefore, information interaction is measured, concerning 4 items (Table 6).

Based on the research of Águeda et al. (2002) [40], and Jaworski et al. (1993) [41], and applications of the scale in recent studies (Aurora et al., 2024) [42], this paper analyzes the actual external business environment faced by logistics service enterprises and CBEC enterprises in Ningbo-Zhoushan port; this study believes that environmental upgrade mainly includes the changes in external conditions affecting technical equipment, service mode, customer demand, and competitor business strategy. Thus, the environmental upgrade is measured through four question items (Table 6).

According to the principle of Balanced Scorecard (BSC), the performance of CBEC enterprises should be considered comprehensively from four dimensions: finance, customer, internal process, and learning growth. This study refers to the scale designed by Stan et al. (1996) [43], and Storey et al. (2001) [44], and applications of the scale in recent studies (Engidaw et al., 2025) [45]; combined with the characteristics of CBEC enterprises, it designs five items from four aspects to measure enterprise performance (Table 6).

Table 6. Variable measures.

Variable	Measure the Item	Reference Frame
PLSI (X)	X1: The enterprise will often put forward the new logistics service concept X2: The enterprise will often put forward the new logistics service technology X3: The enterprise will often put forward a new logistics service business process X4: The enterprise will often put forward a new logistics service management mode	Pim et al. (2001) [33]; Calantone et al. (2002) [34]
Information Interaction (M)	M1: The company often shares the market dynamics with its customers M2: The company often communicates with customers on business processes M3: The company often discusses innovation and creativity with customers M4: The company often provides value experience feedback to customers	Ennew et al. (1999) [36]; Yen et al. (2004) [37]; McEvily (2005) [38]
Environment Upgrade (Z)	Z1: The technology and equipment of the industry where the enterprise is located change rapidly Z2: The service mode of the industry where the enterprise is located changes quickly Z3: The customer demand of the industry where the enterprise is located changes quickly Z4: The business strategy of the competitors in the industry changes quickly	Águeda (2002) [40]; Jaworski (1993) [41]
CBEC Enterprise Performance (Y)	Y1: The service innovation of logistics enterprises has improved the return on investment Y2: The service innovation of logistics enterprises has improved the market share of enterprises in this period Y3: The service innovation of logistics enterprises has improved their customer satisfaction Y4: The service innovation of logistics enterprises enhances the learning and growth potential of the enterprise Y5: The service innovation of logistics enterprises optimizes the internal business process of the enterprises	Stan et al. (1996) [43]; Storey et al. (2001) [44]

In this study, the variables were measured using the Likert 5-point scale: 1 represents "very disagree", 2 means "more disagree", 3 means "general", 4 represents "more consent", and 5 represents "very agree".

3.5. Data Verification

3.5.1. Homology Deviation

As for the questionnaire data, since the respondents are completed by one person in the process of filling in the questionnaire, this may lead to common method variance (CMV). As a result, in terms of procedural control, this study adopted clear and concise items as much as possible, disrupted the order of the questions, reduced the respondent's guess of the measurement purpose, and used anonymous methods when filling in the questionnaire. For statistical control, this study firstly used the Harman single-factorial method; the results explained 74.595% of the total variation, and principal component 1 explained 28.794% of the variance, and 38.600% of the cumulative variance, and did not exceed 40% of the total interpretation (Table 7). Secondly, the confirmatory factor analysis (CFA) is used to analyze all the items in one factor. The fitting indicators of the model are shown in Table 8: chi-square degrees of freedom X^2/df is 4.397 greater than 3, the approximation error root mean square RMSEA is 0.126 greater than 0.1, and the Tucker–Lewis index TLI is 0.804 less than 0.9. The comparative fit index CFI is 0.853 less than 0.9. None of the fit indicators reached the standard, indicating poor model fit, indicating that the study scale data cannot be focused into a factor; that is, there is no serious common method bias in the study data.

Table 7. Homology tests.

Ingredient	The Sum of the Rotating Load Squares		
	Characteristic Root	Variance Interpretation Rate %	Accumulate %
1	8.466	28.794	28.794
2	1.926	20.338	49.132
3	1.376	20.090	69.222
4	0.914	5.373	74.595
Accumulated Variance Interpretation of The Principal Component 1 Occupancy Factor			38.600

Table 8. Single-factor model fit indicators.

Model	X^2	df	X^2/df	RMSEA	TLI	CFI
Unifactor Model	426.465	97	4.397	0.126	0.804	0.853
Requirement	-	-	<3	<0.10	>0.9	>0.9

Note: $X^2/df < 3$ indicates a good model fit; RMSEA < 0.10 indicates a good model fit; TLI > 0.9 indicates a good model fit; CFI > 0.9 is considered a good model fit.

3.5.2. Reliability and Validity Analysis

This study used SPSS 26.0 to test the reliability and validity of the questionnaire data. Reliability indicates the reliability, consistency, and stability of the measurements, usually measured using the Cronbach's α coefficient. The value of this coefficient is between 0 and 1: when the value is above 0.9, the study reliability is excellent; between the interval [0.7 and 0.9], the study reliability is high; and, below 0.7, 0 is selected as the criterion for determining the reliability. The calculation results show that the Cronbach's α coefficient values of PLSI, information interaction, environmental upgrade, and CBEC enterprise performance are all greater than 0.7, indicating that the questionnaire of this study has good reliability (Table 9).

Table 9. Reliability and validity analysis.

Variable	Item	Factor Loading	Cronbach's α	CR	AVE	(X)	(M)	(Z)	(Y)
PLSI (X)	X1	0.720	0.911	0.877	0.641	0.801			
	X2	0.825							
	X3	0.848							
	X4	0.803							
Information Interaction (M)	M1	0.698	0.895	0.806	0.510	0.639 **	0.714		
	M2	0.727							
	M3	0.715							
	M4	0.716							
Environmental Upgrade (Z)	Z1	0.815	0.876	0.834	0.557	0.586 **	0.710 **	0.746	
	Z2	0.694							
	Z3	0.777							
	Z4	0.692							
CBEC Enterprise Performance (Y)	Y1	0.690	0.873	0.877	0.591	0.477 **	0.499 **	0.520 **	0.769
	Y2	0.783							
	Y3	0.656							
	Y4	0.867							
	Y5	0.826							

Note: ** for $p < 0.01$.

Validity mainly indicates the accuracy of the measurement indicators, which is divided into content validity, convergence validity, and differentiation validity. KMO and Bartlett sphere tests yielded a KMO value of 0.877 greater than 0.7 and a significance $p < 0.001$, indicating that the questionnaire data were available for factor analysis (Table 10). In this study, the confirmatory factor analysis of the model showed that the standardized factor load was greater than 0.6, indicating that the combined reliability CR of each factor was greater than 0.7, and the average extraction variance AVE of all constructs was above 0.5, indicating that the square root of the average extraction variance AVE of each factor was greater than the absolute value of the correlation coefficient between this factor and other factors, indicating that the study had good discriminatory validity (Table 9). In conclusion, the variable scales in this study have high reliability and validity.

Table 10. KMO and Bartlett Sphere Test.

KMO	Bartlett Sphere Test		
	Approximate Chi-Square	Degree of Freedom	Significance
0.877	3997.242	136	0.000

4. Empirical Analysis

4.1. Descriptive Statistical Analysis and Correlation Analysis

The mean value and standard deviation of the four variables, PLSI, information interaction, environmental upgrade, and CBEC enterprise performance, as well as the Pearson correlation coefficients between the variables, are shown in Table 11. The results show that the correlation coefficients of all variables are all significant at the 0.01 level (two-tailed), and the PLSI, information interaction, and environment upgrade are all significantly and positively correlated with the performance of CBEC enterprises. This provides support for the validation of H1.

Table 11. Descriptive statistics and correlation analysis (N = 310).

Variable	Mean	Standard Error	(X)	(M)	(Z)	(Y)
PLSI (X)	4.104	0.785	1			
Information Interaction (M)	4.052	0.743	0.639 **	1		
Environment Upgrade (Z)	3.923	0.792	0.585 **	0.710 **	1	
CBEC Enterprise Performance (Y)	4.093	0.644	0.477 **	0.499 **	0.520 **	1

Note: ** for $p < 0.01$.

4.2. Collinearity Test

The regression results may be affected by the "high overlap or similarity between variables in the regression model"; that is, the conclusion is distorted due to the presence of multicollinearity between variables. Therefore, before testing the research hypothesis, it is also necessary to test whether there is multicollinearity between the variables. In this study, a multicollinearity test was conducted through SPSS26.0, and the results are shown in Table 12.

Table 12. Collinearity test.

	B	Standard Error	β	t	Tolerance	VIF
(Constant)	1.917	0.183		10.455		
PLSI (X)	0.170	0.051	0.207 ***	3.301	0.558	1.791
Information Interaction (M)	0.145	0.063	0.167 **	2.285	0.413	2.422
Environmental Upgrade (Z)	0.227	0.056	0.279 ***	4.031	0.458	2.184

Note: *** for $p < 0.001$,** for $p < 0.01$.

When judging multicollinearity, it can be identified by observing the value of tolerance or VIF (tolerance and VIF are reciprocal). The higher the VIF value, the more serious the collinearity, and, when the VIF value is greater than 10, it indicates that there is a serious multicollinearity problem, and there are many overlapping parts between variables, which will affect the authenticity of the conclusion. The tolerance is between 0 and 1, and, when it is less than or equal to 0.1, it also indicates that there is a collinearity problem between variables. As can be seen from Table 12, the maximum value of VIF is 2.422, much less than 10, indicating that there is no multicollinearity problem among the variables in this study.

4.3. Hypothesis Test

4.3.1. Main Effect Test

This study uses SPSS26.0 to test the main effects of PLSI, information interaction, and CBEC enterprise performance. Taking PLSI as the predictive variable, and CBEC enterprise performance and information interaction as the outcome variable, with information interaction as the predictive variable and CBEC enterprise performance as the result variable, regression analysis was conducted, respectively, and the results were shown in Table 13.

Table 13. Main effects test of PLSI, information interaction, and CBEC enterprise performance (N = 310).

Predictor Variable	Outcome Variable	R^2	Adj R^2	F	β	t
PLSI (X)	CBEC Enterprise Performance (Y)	0.228	0.225	90.845	0.392 ***	14.475
PLSI (X)	Information Interaction (M)	0.408	0.406	212.155	0.605 ***	14.566
Information Interaction (M)	CBEC Enterprise Performance (Y)	0.249	0.247	102.290	0.433 ***	10.114

Note: *** indicates $p < 0.001$.

The results in Table 13 show that the standardized regression coefficients of PLSI on CBEC enterprise performance, PLSI on information interaction, and information interaction on CBEC enterprise performance are 0.392, 0.605, and 0.433, respectively, and the regression coefficients are significant at the 0.001 level. The results show that PLSI has a positive impact on the performance of CBEC enterprises, and the correlation degree is significant. PLSI has a positive impact on information interaction, and the correlation degree is significant. Information interaction has a positive impact on the performance of CBEC enterprises, and the correlation degree is significant. Therefore, H1a, H1b, and H1c are assumed to be valid.

4.3.2. Mediating Effect Test

The bootstrap method was adopted in this study; it tested the mediating effect of information interaction between PLSI and CBEC enterprise performance, and used the SPSS macro program Process 3.3 to achieve this. The principle of the Bootstrap method is to extract again and again, obtain the estimated value obtained by multiplying coefficients, and take 50% as the point for estimation to obtain the point estimate. The 95% confidence interval of the point estimate can be formed if the estimated value ranges from 2.5% to 97.5%. If 0 occurs within this range, it means that the mediating effect does not exist. If there is no 0, the mediating effect exists.

Based on Model 4 in the SPSS software, this paper tests the mediating effect of information interaction between PLSI and CBEC enterprise performance. Table 14 shows that PLSI has a significant impact on the performance of CBEC enterprises (β = 0.392, t = 9.531, and $p < 0.001$), and, when the mediating variable is added, PLSI still has a significant impact on the performance of CBEC enterprises (β = 0.220, t = 4.285, and $p < 0.001$). PLSI can positively predict information interaction and have a significant impact (β = 0.605, t = 14.566, and $p < 0.001$), and information interaction can positively predict the performance of CBEC enterprises and have a significant impact (β = 0.285, t = 5.263, and $p < 0.001$).

Table 14. Test of the mediating effect of information interaction (N =310).

Outcome Variable	Predictor Variable	R	R2	F (df)	β	t
CBEC Enterprise Performance (Y)	PLSI (X)	0.477	0.228	90.845	0.392 ***	9.531
Information Interaction (M)	PLSI (X)	0.639	0.408	212.156	0.605 ***	14.566
CBEC Enterprise Performance (Y)	PLSI (X)	0.540	0.292	63.207	0.220 ***	4.285
	Information Interaction (M)				0.285 ***	5.263

Note: *** indicates $p < 0.001$.

As shown in Table 15, the Bootstrap 95% confidence interval value of the direct effect formed by the correlation between PLSI and CBEC enterprise performance ranges from 0.119 to 0.320, without 0, and the Bootstrap 95% confidence interval value of the information-interaction-mediating effect ranges from 0.088 to 0.270. There is no zero. PLSI can directly predict the performance of CBEC enterprises, on the one hand, and indirectly predict the performance of CBEC enterprises through the mediating role of information interaction, on the other hand. The direct effect (0.220) and the mediating effect (0.172)

accounted for 56.122% and 43.878% of the total effect (0.392), respectively. Thus, hypothesis H2 is confirmed.

Table 15. The breakdown of total effect, mediating effect, and direct effect.

	Effect Value	Boot Standard Error	Boot CI Lower Limit	Boot CI Superior Limit	Opposite Effect Value
Gross Effect	0.392	0.041	0.311	0.473	
Direct Effect	0.220	0.051	0.119	0.320	56.122%
The Mediating Effect of Information Interaction	0.172	0.046	0.088	0.270	43.878%

4.3.3. Moderating Effect Test

This study first standardized the variables of PLSI, information interaction, and environmental upgrade, and then multiplied the information interaction and environmental upgrade to obtain the interaction term of the two, to test the moderating effect. The results are shown in Table 16: After adding environmental upgrade into the model, the product term of information interaction and environmental upgrade has a significant predictive effect on the performance of CBEC enterprises ($\beta = 0.139$, $t = 2.944$, $p < 0.01$), indicating that environmental upgrade positively moderates the predictive relationship between information interaction and the performance of CBEC enterprises.

Table 16. Test of the moderating effect of environmental upgrade (N =310).

Outcome Variable	Predictor Variable	R	R^2	F (df)	β	t
CBEC Enterprise Performance (Y)		0.588	0.346	40.335		
	PLSI (X)				0.145 **	2.822
	Information Interaction (M)				0.180 **	2.833
	Environment Upgrade (Z)				0.212 **	3.785
	Information Interaction × Environment Upgrade				0.139 **	2.944

Note: ** indicates $p < 0.01$.

To further verify the interaction between information interaction and environment upgrade, the values obtained by adding or subtracting 1 standard deviation from the mean value of environment upgrade were divided into the "high environment upgrade group" and "low environment upgrade group", and then a simple slope analysis was performed. The moderating effect is shown in Figure 8.

It can be seen from Figure 8 that, under different environmental upgrade levels, information interaction has different slopes on the performance of CBEC enterprises: when the level of environmental upgrade is high (M + 1 SD), the slope of the straight line (simple slope = 0.290) is larger than that (simple slope = 0.071) when the level of environmental upgrade is low (M − 1 SD), which indicates that information interaction has a stronger positive impact on the performance of CBEC enterprises. This further verifies that the improvement of the level of environmental upgrade strengthens the promoting effect of information interaction on the performance of CBEC enterprises. Hypothesis H3 is verified.

4.3.4. Moderated Mediating Effect Test

To test whether there is a moderated mediating effect in this study, the Process plug-in was used to integrate relevant variables into a Model for verification and analysis. A Model 14 consistent with this study was selected, and the sample size of Bootstrap was set to
5000 with a confidence interval of 95%. The specific operation results are shown in Table 17.

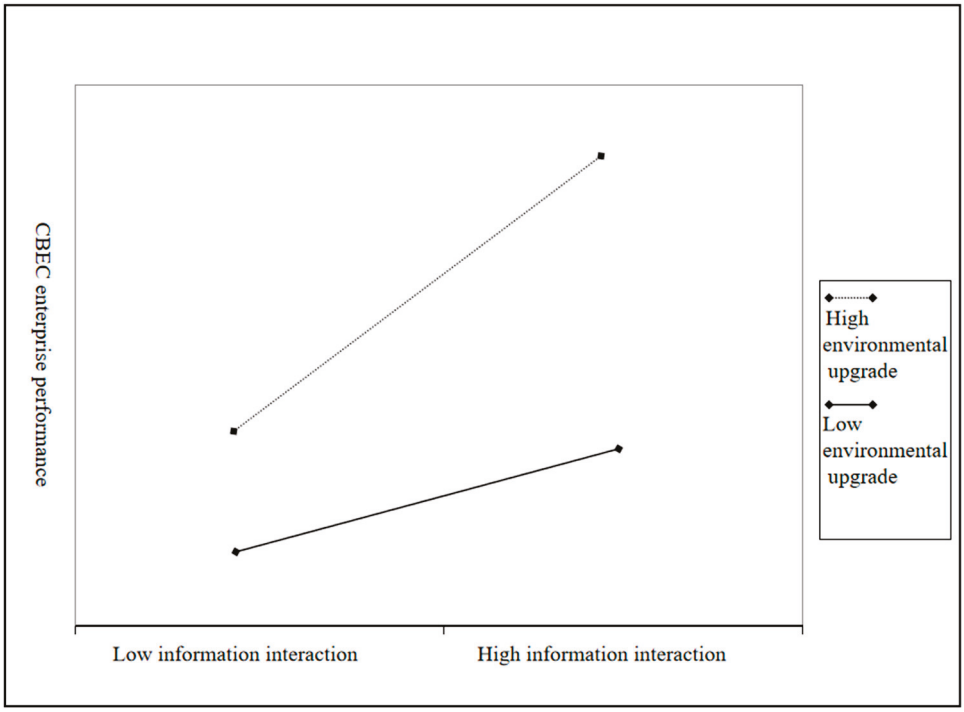


Figure 8. The moderating effect of environmental upgrade.

Table 17. The mediating effect of information interaction under different environment upgrade levels.

Independent Variable	Moderator Variable	Indirect Effect	Standard Error	95% Confidence Interval	
Information Interaction (M)	Low Environmental Upgrade (Z)	0.043	0.060	−0.060	0.181
	High Environmental Upgrade (Z)	0.175 ***	0.066	0.047	0.304
	The Moderated Mediated Index	0.084 **	0.048	0.034	0.156

Note: High environment upgrade and low environment upgrade, respectively, mean plus and minus one standard deviation; *** means $p < 0.001$, ** means $p < 0.01$.

It can be seen from Table 17 that environmental upgrade has a significant moderating effect on the mediating effect of information interaction between PLSI and CBEC enterprise performance (moderating index is 0.084, $p < 0.01$, 95% confidence interval [0.034, 0.156], excluding 0), indicating that there is a moderated mediating effect.

When the environment upgrade level is low (M − 1 SD), the 95% confidence interval [−0.060, 0.181] contains 0, indicating that the mediating effect of information interaction is not significant at the low environment upgrade (M − 1 SD). When the level of environment upgrade is high (M + 1 SD), the 95% confidence interval [0.047, 0.304] does not contain 0, indicating that the mediating effect of information interaction is significant. Moreover, it can be seen from Table 17 that the indirect effect value (0.175) of the high environment upgrade (M + 1 SD) is larger than that of the low environment upgrade (M−1 SD) (0.043), which further indicates the moderating effect of environment upgrade on the mediating effect of information interaction. In other words, the mediating effect of information interaction between PLSI and CBEC enterprise performance is moderated by environmental upgrades. The higher the level of environmental upgrade, the stronger the mediating effect of information interaction, and H4 has been verified.

5. Conclusions and Future Prospects

5.1. Research Conclusions

This study focuses on port logistics service, focuses on the intermediary role of information interaction and the regulating role of the environmental upgrade, demonstrates the function mechanism of PLSI on the performance of CBEC enterprises, and obtains the following research conclusions:

First, PLSI has a significant and positive impact on the performance of CBEC enterprises. As cross-border e-commerce has gradually become a new growth pole of China's import and export, its customer needs are becoming more and more personalized, and the demand for customized services in the industry is increasing. As a transportation hub on the cross-border logistics chain, port logistics enterprises will respond in a timely manner to the diversified service needs of CBEC enterprises by providing new service modes, service technologies, and equipment, which will directly promote the performance improvement of CBEC enterprises.

Second, PLSI has a significant positive impact on information interaction. Customer participation has an important impact on enterprise innovation. In the process of innovating and reforming logistics services, port logistics enterprises cannot leave the customer demand orientation of CBEC enterprises, which will promote the improvement of the frequency and expansion of the scope of information exchange between the two. Port logistics enterprises draw from the reform of reform to make the PLSI more accurate and efficient.

Third, information interaction has a significant positive impact on the performance of CBEC enterprises. Information exchange and communication with port logistics enterprises enable CBEC enterprises to deeply participate in the development process of new modes and processes of port logistics services, create "tailor-made" logistics service technologies and scenarios, further compress logistics time, optimize logistics nodes, and greatly improve the efficiency and business performance of cross-border logistics.

Fourth, information interaction plays an intermediary role between PLSI and the performance of CBEC enterprises. CBEC enterprises are the core of port logistics enterprises. It is the fundamental goal to fully tap their potential customer needs and continuously innovate to provide a higher-level cross-border logistics service experience. It is an important way to generate logistics service innovation by strengthening interpersonal interaction with CBEC enterprises to promote information sharing, and to obtain more information about customer preferences, process evaluation, and service feedback.

Fifth, whether the real business performance improvement of the latter can be realized through the information interaction between port logistics enterprises and CBEC enterprises is also regulated by the environmental upgrade. The external environment upgrade, the faster uncertainty of the enterprise management, the two sides for information sharing, awareness of valuable information, and the effectiveness of the information identification use are higher and will master the dominant and recessive information into PLSI and increase the chance of CBEC performance being higher.

5.2. Theoretical Significance

First, this study breaks through the existing literature, which mainly focuses on the port logistics service itself's internal innovation mechanism, relatively ignoring the service innovation to the service object of the external spillover mechanism of limitation. This study takes port logistics as the breakthrough point, not only enriching the service's leading logic theory and value theory, but also broadening the service innovation research boundary, and deepening the port logistics service which affects the CBEC "black box" theory, from the perspective of logistics service innovation.

Second, this study brings in the key process variable information interaction. Specifically, this study analyzes the information interaction in the PLSI and CBEC enterprise performance between intermediary mechanisms, and expounds the information resource sharing for the important role of PLSI, to improve the cross-border logistics chain port logistics enterprises and CBEC business effect between the causal chain theory explanation, and clear the PLSI of CBEC enterprise performance of the micro-mechanism.

Third, this study integrates the key external factor environment upgrade. That is, this study builds a regulated mediation model, to explore the environment upgrade as an important boundary condition, and how to regulate the mediation mechanism of information interaction, to refine the PLSI through information interaction on how CBEC enterprise performance influences the role of a path, further broadening the main effect of extension boundaries, complementing and enriching the existing theory.

5.3. Empirical Significance for Management Enlightenment

First, we should increase the investment of innovation resources and pay attention to the innovation of logistics services. With the advent of the service economy era, the role of the customer is in continuous evolution, from past passive consumers to customers who are now focused on the experience of feeling. The personalized, customized demands of CBEC enterprises constantly emerge. Port logistics enterprises need to increase innovation resources and speed up the innovation logistics service for a timely response to the new needs of customers. First, with regard to upgrading professional hardware and software resources, the efficiency of port logistics service should be improved by updating the storage, transportation, loading, and unloading equipment, optimizing the operation process, and using advanced information management technology. Secondly, with regard to introducing professional talent resources, it is essential that we accelerate the gathering of more port compound talents, optimize the structure of professional talents, improve the team innovation ability, and break through the frontier service innovation bottleneck of port logistics. Thirdly, strengthening the industry–university–research cooperation mechanism. We will promote communication with universities and scientific research institutions, vigorously carry out the construction of various joint construction platforms and research on related projects, actively promote the transformation of innovation achievements, and create a good atmosphere for innovation. Fourthly, with regard to strengthening the innovation consciousness and innovation ability, as the main body of enterprise logistics service innovation, whether the employees have innovation consciousness and innovation motivation determines whether they will produce innovative behavior. Port logistics enterprises should not only strengthen the employees' understanding of innovation through training, but also improve the innovation incentive system of the employees, actively seek the thrust and pull that influence the logistics service innovation, and maintain the employees' innovation passion to the greatest extent.

Second, we should strengthen the construction of information collection and acquisition and co-creation and utilization mechanisms, and pay attention to filling the information blind area. In the era of big data where information, data, and other intangible assets such as information and data generally become key factors of production and operation, port logistics enterprises and CBEC enterprises must pay more attention to smooth information communication and feedback. Adequate and high-quality information is an important source of enterprise innovation and creativity. Firstly, we should build a convenient and friendly communication environment. Enterprises should actively create a good atmosphere of mutual trust for information exchange, and open mutual emotional links, and ensure that information sharing is frank and sufficient. Secondly, we should build a perfect new media information exchange platform. Based on the characteristics of fast commu-

nication speed, wide coverage, strong interactivity, traceability, and diversification, we actively improve communication carriers with customers such as we-media, social media, and mobile media, to realize knowledge sharing and continuously generate core competitiveness. In addition, we should vigorously advocate for the timely tracking experience of enterprises and active dialogue and reflection between enterprises and consumers, to promote the deep participation of various partners in the cross-border logistics chain and realize the co-creation value in a real sense.

Third, we should speed up the dynamic assessment of the external environment and the optimization of the rapid response capacity, and pay attention to the cultivation of business resilience. Port logistics enterprises must have strong operational resilience to seize the market opportunity in the uncertain environment and achieve sustainable development. Business resilience is a kind of internal invisible ability for enterprises to effectively respond to the impact of emergencies in a dynamic environment. Enterprises need to identify and respond to the opportunities and challenges brought by the dynamic environment. It is the key business ability that enterprises need to attach great importance to when the external environment tends to be dynamic and uncertain. First, the ability to identify sensitive environment upgrades should be established. To observe and interpret technology equipment upgrades, customer demand changes, the evolution of market structure, policy guidance adjustment of external events to specific departments, and personnel, as an important daily work content continues to advance, for major changes, should be established by the daily recognition of individual evaluation monitoring mechanisms. Second, build up the ability to effectively respond to environmental shocks. The external environment upgrade is uncontrollable. Port logistics enterprises should predict the future, do a good job of "scenario planning", prepare for a rainy day, formulate targeted response measures in advance, and invite experts and scholars to study and plan jointly when necessary. In addition, we will actively seek the support of the government. Under the background of the national advocacy for the construction of digital ports and smart logistics, port logistics enterprises should take advantage of the situation to strive for policy support, financial subsidies, tax incentives, etc., rely on government support to accelerate the innovation of port logistics services, enhance operational resilience, and effectively resist the negative effects brought by environmental impact.

5.4. Research Limitations and Outlook

This study significantly contributes to enriching relevant theoretical frameworks and broadening the research horizons within its domain; it is imperative to acknowledge certain inherent limitations. Firstly, the sample enterprise diversity is constrained. Although the study distinguishes port logistics enterprises into the warehousing, transportation, and comprehensive categories, CBEC enterprises remain unclassified in a comparable manner, resulting in an under-representation of the latter. To address this, future endeavors should meticulously categorize CBEC enterprises based on their unique characteristics and varied transaction modalities, thereby investigating whether PLSI exerts differential impacts across these diverse enterprises. At the same time, the sample size can be expanded, and the conclusion of this study can be verified repeatedly by investigating the top management of enterprises to understand the enterprises, so as to enhance the reliability of the research conclusion.

Secondly, the geographical scope of the investigated enterprises is limited. Rooted in the analysis of logistics and CBEC enterprises operating within Ningbo-Zhoushan port, the study fails to capture the intricate interplay between port logistics service innovation and the multifaceted factors that encompass natural port conditions, technological advancements, local government policies, and other regional nuances, all of which can significantly

influence innovation efficiency and outcomes. To mitigate this, future research should expand its geographical reach, ensuring a balanced representation of samples across regions, and subsequently differentiate and compare the findings based on regional specificities.

Lastly, the study's examination of endogeneity constraints leaves room for improvement. Specifically, there remains a need to delve into the potential reverse causality between PLSI and CBEC enterprise performance, neglecting the intricate endogenous relationships among variables. To strengthen this aspect, future investigations could incorporate lagged terms into regression models, and employ linear feedback tests or leverage structural equation modeling to not only extend the research frontier but also deepen our understanding of the intricate dynamics at play.

Author Contributions: Conceptualization, W.J.; methodology, W.J. and H.L.; validation, W.J. and Z.W.; resources, Y.J. and W.J.; writing and editing, W.J. and Y.J. All authors have read and agreed to the published version of the manuscript.

Funding: The Key Research Center of Philosophy and Social Science of Zhejiang Province—Modern Port Service Industry and Creative Culture Research Center (2022JDKTYB41), General Scientific Research Fund of Zhejiang Provincial Education Department (Y202249631), and General Fund Project of Ningbo National Science Foundation (20221JCGY010743).

Institutional Review Board Statement: Ethical approval was not required for this study.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data is available; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lin, C.; Li, C. Analysis of Key Influential Factors of Chinese E-Commerce in Africa Based on Swot-Pest Analysis Model Empirical Evidence. *World J. Manag. Sci.* **2024**, *2*, 7–23.
2. Natalia, K.; Anastasia, G. Digital Platform for Maritime Port Ecosystem: Port of Hamburg Case. *Transp. Res. Procedia* **2021**, *54*, 909–917. [CrossRef]
3. Christopher, M.; Laima, G.; Lawrence, H. Cross-Border Capacity-Building for Port Ecosystems in Small and Medium-Sized Baltic Ports. *Balt. J. Eur. Stud.* **2021**, *11*, 113–132.
4. Luis, R.; Matthias, D. The Uneven Role of Users in Service Innovation Performance. *Econ. Innov. New Technol.* **2023**, *32*, 953–976.
5. Sipho, S.M. Towards a Conceptual Model of Measuring the Influence of Service Innovation on Business Performance. *Int. J. Multidiscip. Bus. Sci.* **2023**, *9*, 24. [CrossRef]
6. Jessica, O.O.; Chinekwu, S.O. Big Data Analytics and AI for Optimizing Supply Chain Sustainability and Reducing Greenhouse Gas Emissions in Logistics and Transportation. *Int. J. Multidiscip. Res. Growth Eval.* **2024**, *5*, 1536–1548.
7. Juying, Z.; Jiehui, L. Policy involvement and policy consistency identification of supportive policies for SMEs. *Int. Entrep. Manag. J.* **2024**, *20*, 2901–2937. [CrossRef]
8. Tao, Z.; Shuliang, Z. Triple helix relationship research on China's regional university–industry–government collaborative innovation: Based on provincial patent data. *Growth Change* **2021**, *52*, 1361–1386.
9. Sedat, B.; Hercules, H. Port competitiveness: Do container terminal operators and liner shipping companies see eye to eye? *Mar. Policy* **2022**, *135*, 104866. [CrossRef]
10. Song, M.J.; Lee, H.Y. The relationship between international trade and logistics performance: A focus on the South Korean industrial sector. *Res. Transp. Bus. Manag.* **2022**, *44*, 100786. [CrossRef]
11. Tien, M.P.; Vinh, V.T. Port service quality (PSQ) and customer satisfaction: An exploratory study of container ports in Vietnam. *Marit. Bus. Rev.* **2021**, *6*, 72–94.
12. Khaled, H.; Dong-Wook, S. Port supply chain integration and sustainability: A resource-based view. *Int. J. Logist. Manag.* **2024**, *35*, 504–530.
13. Giuseppe, M.; Antonio, C.; Giovanna, C. Evaluation of structural factors in a third-generation port: Methods and applications. *Int. J. Transp. Dev. Integr.* **2022**, *6*, 347–362.

14. Theo, N.; Hercules, H. The Red Sea Crisis: Ramifications for vessel operations, shipping networks, and maritime supply chains. *Marit. Econ. Logist.* **2024**, *26*, 1–20. [CrossRef]
15. Stephen, L.V.; Robert, F.L. Institutions and Axioms: An Extension and Update of Service-Dominant Logic. *J. Acad. Mark. Sci.* **2016**, *44*, 5–23.
16. Zeeshan, R.; Johan, W.; Ceren, A.V. Digital transformation of maritime logistics: Exploring trends in the liner shipping segment. *Comput. Ind.* **2023**, *145*, 103811.
17. Yasin, S.; Raymond, L. Characterizing the spaces of consumer value experience in value co-creation and value co-destruction. *Eur. J. Mark.* **2022**, *56*, 105–136. [CrossRef]
18. Motshedisi, S.M.; Willie, T.C. Value co-creation as a mediator between strategic planning and social enterprise performance. *Soc. Enterp. J.* **2023**, *19*, 23–39.
19. Daniel, A.L.; Brian, W. Resource Redeployment and the Pursuit of the New Best Use: Economic Logic and Organizational Challenges. *Strategy Sci.* **2024**, *10*, 32–47.
20. Sunday, B. A Complementarity Perspective of Knowledge Resources. *J. Knowl. Econ.* **2022**, *13*, 1300–1320.
21. Dan, J. Service dyads: Understanding the dynamics of social perspective-taking and value co-creation amidst service problems. *Serv. Bus.* **2024**, *18*, 287–313.
22. Divesh, O.; Mumin, D. Social exchange in buyer-supplier relationships and innovation speed: The mediating and moderating role of information sharing and knowledge channels. *J. Knowl. Manag.* **2023**, *27*, 1509–1533.
23. Ann, B.; Simon, A. *Interacting with Information*; Springer Nature: Berlin, Germany, 2022.
24. Saqib, M.; Samera, N. Achieving supply chain sustainability: Enhancing supply chain resilience, organizational performance, innovation and information sharing: Empirical evidence from Chinese SMEs. *Mod. Supply Chain Res. Appl.* **2025**, *7*, 2–29.
25. Tobias, M.; Andreas, S. The Effect of Work Ethic on Employees Individual Innovation Behavior. *Creat. Innov. Manag.* **2017**, *26*, 391–406.
26. Wang, Q. Path to Enhance the International Competitiveness of Cross-Border E-Commerce Logistics Enterprises-Research Based on AHP Analysis Method. *Int. J. Front. Sociol.* **2022**, *4*, 37–45.
27. Yafei, T.; Zhaohui, Z. The effect of ESG rating events on corporate green innovation in China: The mediating role of financial constraints and managers' environmental awareness. *Technol. Soc.* **2022**, *68*, 101906. [CrossRef]
28. Sirmon, D.G.; Hitt, M.A. Managing firm resources in dynamic environments to create value: Looking inside the black box. *Acad. Manag. Rev.* **2007**, *32*, 273–292. [CrossRef]
29. Aprisma, R.; Sudaryati, E. Environmental uncertainty and firm performance: The moderating role of corporate governance. *J. Akunt.* **2020**, *24*, 87–203. [CrossRef]
30. Sprong, N.; Driessen, P.H.; Hillebrand, B. Market innovation: A literature review and new research directions. *J. Bus. Res.* **2021**, *123*, 450–462. [CrossRef]
31. Kamkankaew, P.; Phattarowas, V.; Khumwongpin, S. Increasing Competitive Environment Dynamics and the Need of Hyper-Competition for Businesses. *Int. J. Sociol. Anthropol. Sci. Rev.* **2022**, *2*, 9–20.
32. Lv, C.L.; Peng, C.; Li, R.X.; Yin, J.Y. The impact of organizational dual learning and its complementarity on firm sustainability performance in a dynamic environment: The mediating role of sustainable innovation capability. *Sci. Technol. Manag. Res.* **2021**, *41*, 135–144.
33. Pim, D.H.; Rob, B. *Services and the Knowledge-Based Economy*; Taylor & Francis Group: Abingdon-on-Thames, UK, 2001.
34. Calantone, R.J.; Cavusgil, S.T.; Zhao, Y. Learning orientation, firm innovation capability, and firm performance. *Ind. Mark. Manag.* **2002**, *31*, 515–524. [CrossRef]
35. Jing, D.; Wen, C.; Jia, J.L.; Yongyi, S. Service innovation of cold chain logistics service providers: A multiple-case study in China. *Ind. Mark. Manag.* **2020**, *89*, 143–156. [CrossRef]
36. Ennew, C.T.; Binks, M.R. Impact of participative service relationships on quality, satisfaction and retention: An exploratory study. *J. Bus. Res.* **1999**, *46*, 121–132. [CrossRef]
37. Yen, H.R.; Gwinner, K.P.; Su, W. The impact of customer participation and service expectation on Locus attributions following service failure. *Int. J. Serv. Ind. Manag.* **2004**, *15*, 7–26. [CrossRef]
38. McEvily, B.; Marcus, A. Embedded ties and the acquisition of competitive capabilities. *Strateg. Manag. J.* **2005**, *26*, 1033–1055. [CrossRef]
39. Bin, W.; Yukun, L.; Sharon, K. How does the use of information communication technology affect individuals? A work design perspective. *Acad. Manag. Ann.* **2020**, *14*, 695–725. [CrossRef]
40. Águeda, E.; Ángel, M.; Molina, A. Market orientation in service: A review and analysis. *Eur. J. Mark.* **2002**, *36*, 1003–1021.
41. Jaworski, B.J.; Kohli, A.K. Market orientation: Antecedents and consequences. *J. Mark.* **1993**, *57*, 53–70. [CrossRef]
42. Aurora, G.M.; Rodrigo, M.R. The key role of innovation and organizational resilience in improving business performance: A mixed-methods approach. *Int. J. Inf. Manag.* **2024**, *77*, 102777. [CrossRef]

43. Stan, B.; Joan, B. Performance measurement in service businesses revisited. *Int. J. Serv. Ind. Manag.* **1996**, *7*, 6–31. [CrossRef]
44. Storey, C.; Kelly, D. Measuring the performance of new service development activities. *Serv. Ind. J.* **2001**, *21*, 71–90. [CrossRef]
45. Engidaw, A.E.; Haichun, Y.; Zou, W. Opportunities and challenges in cross-border e-commerce: Strategic management within the legal context of BRI countries—A systematic literature synthesis and future research directions. *Technol. Anal. Strateg. Manag.* **2025**. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Spatial Heterogeneity of Intra-Urban E-Commerce Demand and Its Retail-Delivery Interactions: Evidence from Waybill Big Data

Yunnan Cai †, Jiangmin Chen † and Shijie Li *

School of Architecture and Urban Planning, Guangdong University of Technology, Guangzhou 510090, China; caiyunnan@gdut.edu.cn (Y.C.); 2112310024@mail2.gdut.edu.cn (J.C.)
* Correspondence: lishijie@gdut.edu.cn
† These authors contributed equally to this work.

Abstract: E-commerce growth has reshaped consumer behavior and retail services, driving parcel demand and challenging last-mile logistics. Existing research predominantly relies on survey data and global regression models that overlook intra-urban spatial heterogeneity in shopping behaviors. This study bridges this gap by analyzing e-commerce demand's spatial distribution from a retail service perspective, identifying key drivers, and evaluating implications for omnichannel strategies and logistics. Utilizing waybill big data, spatial analysis, and multiscale geographically weighted regression, we reveal: (1) High-density e-commerce demand areas are predominantly located in central districts, whereas peripheral regions exhibit statistically lower volumes. The spatial distribution pattern of e-commerce demand aligns with the urban development spatial structure. (2) Factors such as population density and education levels significantly influence e-commerce demand. (3) Convenience stores play a dual role as retail service providers and parcel collection points, reinforcing their importance in shaping consumer accessibility and service efficiency, particularly in underserved urban areas. (4) Supermarkets exert a substitution effect on online shopping by offering immediate product availability, highlighting their role in shaping consumer purchasing preferences and retail service strategies. These findings contribute to retail and consumer services research by demonstrating how spatial e-commerce demand patterns reflect consumer shopping preferences, the role of omnichannel retail strategies, and the competitive dynamics between e-commerce and physical retail formats.

Keywords: e-commerce demand; waybill big data; spatial consumer behavior; retail logistics; last-mile delivery; omnichannel retailing

1. Introduction

The exponential growth of online shopping has significantly propelled global e-commerce demand [1]. According to Ti Insight's Global Express & Small Parcels 2024 report, the market size is projected to grow from EUR 565.9 billion in 2024 to EUR 736.6 billion in 2028 [2]. China, the world's largest parcel market, accounted for one-third of the global parcel volume in 2023, with an average of 94 parcels per capita annually [3], and this figure continues to rise. However, this rapid expansion has simultaneously highlighted and exacerbated the vulnerabilities within urban logistics networks [4–6]. The disparity in delivery efficiency between central urban areas and suburbs [7,8] becomes particularly pronounced during peak shopping periods [9]. Retailers face numerous systemic challenges, including sharply escalating delivery costs and the persistent difficulty of ensuring timely deliveries to peripheral urban areas [10]. This spatial inequality in delivery service accessibility is

increasingly emerging as a critical bottleneck constraining the inclusivity and sustainability of e-commerce.

Addressing these challenges necessitates a thorough understanding of the interdependencies among online shopping, e-commerce operations, and parcel delivery. These elements form a tightly coupled feedback loop: online shopping demand drives the functioning of e-commerce systems [11], which in turn generates parcel delivery demand [12]. Conversely, the efficiency and accessibility of delivery services influence consumers' online shopping intentions and behaviors. Within this loop, parcel delivery volume serves as a robust proxy indicator for e-commerce demand [13]. Specifically, the number of parcels delivered reflects the intensity of online shopping activity, as each online transaction typically necessitates a physical delivery. Consequently, waybill data can effectively capture the spatial distribution patterns of e-commerce demand, providing a detailed and objective measure of consumer behavior's spatial manifestations. Nevertheless, traditional research methods, such as questionnaires and linear regression models, often struggle to adequately capture the inherent spatial heterogeneity of e-commerce demand [14–16]. Although recent studies have begun leveraging waybill data and incorporating spatial econometric methods (e.g., Moran's I index [17] and the Spatial Durbin Model [18]), they still fall short in revealing the spatial non-stationarity of key influencing factors—that is, the phenomenon where the magnitude and nature of their effects vary geographically.

In practice, both e-commerce demand and the spatial provision of supporting facilities within cities exhibit pronounced heterogeneity; however, existing theoretical analyses have yet to adopt methods that explicitly capture this spatial non-stationarity. This disjuncture impedes the alignment between supply-side capabilities and demand-side requirements, posing a critical challenge for cities seeking to formulate differentiated and place-based planning policies. To bridge this gap, this study utilizes waybill big data from the 2023 Singles' Day (November 11) shopping festival to accurately map the spatial distribution of e-commerce demand and rigorously examine the geospatial heterogeneity of its underlying drivers. By doing so, this research aims to deepen the understanding of the spatial patterns and mechanisms governing e-commerce demand and provide robust empirical support for developing tailored, place-based logistics infrastructure planning policies capable of effectively meeting diverse regional needs. The remainder of this paper is structured as follows. Section 2 reviews the relevant literature. Section 3 outlines the study area, methodology, and data sources. Section 4 presents and discusses the modeling results. Finally, Section 5 concludes the study.

2. Literature Review

A deep understanding of the spatial heterogeneity of e-commerce demand is crucial for effectively addressing the interconnected challenges of online shopping, e-commerce operations, and parcel delivery services. Current research on e-commerce demand primarily unfolds along two distinct paths: the first path focuses on using questionnaire surveys to directly investigate online shopping demand itself; the second path employs waybill data to indirectly infer delivery demand through reverse reconstruction. However, both approaches face methodological limitations in capturing spatial heterogeneity. Table 1 summarizes the research perspectives, data sources, and analytical methods of relevant studies.

Table 1. Summary of previous studies on e-commerce demand.

Research Perspective	Data Source	Driving Factors	Analytical Methods	Reference
Online shopping behavior	2014 Online Shopping Index (OSI) by Alibaba Group	Socioeconomic: urban administrative level, Internet penetration level, age, educational level, income level Built environment: number of shopping centers, number of buses, number of express delivery points	General Spatial Model (SAC) includes both the spatial lag term and the spatial error term	[19]
Online shopping behavior	Dutch e-commerce survey (1996–2001)	Socioeconomic: gender, age, educational level, Internet penetration level, credit card popularity level, family shopping experience Built environment: urbanization level, accessibility of physical stores	Ordinary Least Squares	[20]
Online shopping behavior	Face-to-face survey in Tel Aviv metropolitan area	Socioeconomic: gender, age, educational level, member of family, income level, the number of cars owned by the family, reasons for Internet shopping, Internet penetration level, enjoy shopping	Ordinal Logit Model	[21]
Online shopping behavior	Survey Monkey's online shopper panel	Socioeconomic: age, educational level, income, number and frequency of mobile purchases, annual mobile purchase expenditure amount	Linear Regression Model	[14]
Online shopping behavior	E-commerce in Belgium 2016 questionnaire by Comeos	Socioeconomic: age, educational level, income Built environment: urbanization level	Logistic Regression Model	[22]
Online shopping behavior	Questionnaires from 384 online shopping platform users	Behavioral: online purchasing habits, loss avoidance, transition costs, uncertainty cost, sunk cost, social norms, information search, evaluation of alternative solutions, negative intent recognition, uncertain purchasing behavior, negative purchasing behavior, negative recommendations	Structural Equation Model	[23]
Online shopping behavior	Online self-administered survey by Oualtrics	Behavioral: product portfolio, product quality, price transparency, website convenience, service quality, security issues, online retail discounts	Partial Least Squares Structural Equation Modeling	[24]
Online shopping behavior	Shopping survey (December 2008–January 2009)	Socioeconomic: age, income, family vehicle ownership status, number of household employees, shopping behavior, Internet experience, shopping attitude Built environment: offline shopping accessibility	Structural Equation Model	[25]

Table 1. *Cont.*

Research Perspective	Data Source	Driving Factors	Analytical Methods	Reference
Online shopping behavior	Acxiom's Research Opinion Poll	Socioeconomic: age, income, family vehicle ownership status, family size, Internet access Built environment: urbanization level, offline shopping accessibility	Binary Logistic Regression Model	[26]
Online shopping behavior	Internet diary survey (April 2003)	Socioeconomic: age, gender, income; educational level, weekly working hours, Internet use level, population density, proportion of white race Built environment: the number of shopping opportunities during travel time, shopping opportunity area during travel time	Logistic Regression Model	[27]
Online shopping behavior	Household survey in Nanjing (July–August 2015)	Socioeconomic: age; income, educational level, usage of smartphones, daily Internet usage time Built environment: distance to the workplace, distance to the nearest subway station, distance to the nearest shopping center	Joint Binary Logit Regression	[28]
Online shopping behavior	Household survey in Shiraz, Iran	Socioeconomic: income, educational level, work situation, driver's license status, offline shopping frequency, online shopping frequency, Internet use experience Built environment: land use diversity, offline shopping and transportation methods, offline shopping locations, proportion of connection nodes, intersection density, employment density, residential density, the distance from home to the nearest store, the distance from home to the nearest bus stop, residential location	Structural Equation Model	[29]
Parcel delivery	Logistics company waybill data (2019–2022)	Socioeconomic: socioeconomic income, nonresident real estate transactions Built environment: density of commercial facilities, diversity of commercial facilities, per capita parcel delivery and opportunities achievable within 15 min under walking distance and socioeconomic conditions, per capita package delivery and opportunities achievable within 30 min under public transportation distance and socioeconomic conditions	Moran's I and bivariate Moran's I	[17]
Parcel delivery	Logistics company waybill data (2019–2021)	Socioeconomic: income, family size Built environment: number of offline retail stores, community area	Negative Binomial Regression Model	[16]

Table 1. *Cont.*

Research Perspective	Data Source	Driving Factors	Analytical Methods	Reference
Parcel delivery	Logistics company waybill data (February–April 2019)	Socioeconomic: population density, average family size, average household income, average age of household heads, average vehicle ownership rate Built environment: average age of residential buildings, accessibility of commercial complexes, accessibility of public transportation	Linear Regression Model	[15]
Parcel delivery	CJ Logistics parcel delivery OD data (June 2014)	Socioeconomic: single occupancy rate, gender, age, daytime population size, GDP Built environment: proportion of residential areas, proportion of commercial areas, apartment ratio, retail area ratio	Spatial Durbin Error Model	[18]

Early research focusing on online shopping demand typically employed questionnaire data, adopting an individual behavioral perspective to directly capture the drivers of online purchasing decisions through consumer self-reports. These studies primarily applied traditional statistical models, such as linear regression [14] and logistic regression [26–28], to analyze socio-economic factors (e.g., gender, age, income, education level) and spatial-contextual factors (e.g., distance to logistics points, in-store shopping experience, accessibility of offline facilities) influencing online shopping demand. Representative works include Farag et al. (2006) applying Ordinary Least Squares (OLS) to Dutch survey data [20]; Rotem-Mindali (2010) using an ordinal logit model with face-to-face survey data from the Tel Aviv metropolitan area [21]; and Beckers et al. (2018) constructing a logistic regression model based on questionnaire data from Comeos, the Belgian retail federation [22]. To explore more complex causal networks among variables, some scholars introduced Structural Equation Modeling (SEM) [23]. For instance, Etminani-Ghasrodashti and Hamidi (2020) integrated household survey data from the Shiraz metropolitan area in Iran, building a causal network model incorporating up to 18 socio-economic and spatial-contextual variables [29]. Additionally, Shao et al. (2022) attempted to combine comprehensive indices constructed by e-commerce platforms and used a Spatial Lag of Error model to explore the influence of virtual and real accessibility [19].

In contrast, research based on waybill data starts from the operational logistics outcomes, reverse-engineering demand characteristics by analyzing actual delivery occurrences. This provides a more objective, inverse proxy measure for gauging e-commerce demand. Studies along this path have utilized diverse analytical methods to investigate the drivers of e-commerce demand: Oliveira et al. (2025) applied Moran's I and bivariate Moran's I to examine spatial autocorrelation in waybill data from Belo Horizonte [17]; de Sousa et al. (2023) employed negative binomial regression on the same dataset [16]; Cheng et al. (2021) applied linear regression models to waybill data [15]; and Shin et al. (2023) utilized a spatial error model for their analysis [18]. Compared to questionnaire-based studies, this line of research places greater emphasis on the integrated analysis of socio-economic factors (e.g., income, population density, education level, household size, GDP) and spatial-contextual factors (e.g., commercial density, retail density, public transport coverage, land-use types, building age).

In summary, while both data sources offer valuable and complementary analytical perspectives, and their methodological approaches differ, they share a fundamental limitation: insufficient characterization of spatial heterogeneity. Traditional regression models, such as linear regression and negative binomial regression, assume global linear relationships between variables, making them ill-equipped to explain the significant spatial variation in the effects of influencing factors across different geographical regions. Although spatial econometric models, such as the spatial error model or spatial Durbin model, can effectively quantify spatial dependence, they still struggle to capture the spatial non-stationarity inherent in key explanatory variables—namely, the phenomenon where the intensity and mode of these factors' impact on e-commerce demand dynamically change according to geographic location.

3. Materials and Methods

3.1. Study Area

Guangzhou is an ideal location for studying e-commerce demand for several reasons. As a major metropolis in Southern China, it boasts a robust and diversified economy, and it covers approximately 7434 square kilometers. This area is organized into three circles: the Inner Ring Expressway, the Ring Road, and the Beltway. According to the Seventh National Census [30], Guangzhou had a resident population of 18.67 million. This makes it one of the most populous and economically significant cities in China, which in turn influences e-commerce demand patterns. Its strategic position in the Pearl River Delta—a hub of economic activity in China—also increases its importance. Guangzhou serves as a crucial node in both national and international trade networks, having excellent land, sea, and air connectivity, and it is well positioned near to Hong Kong and Macau (see Figure 1).

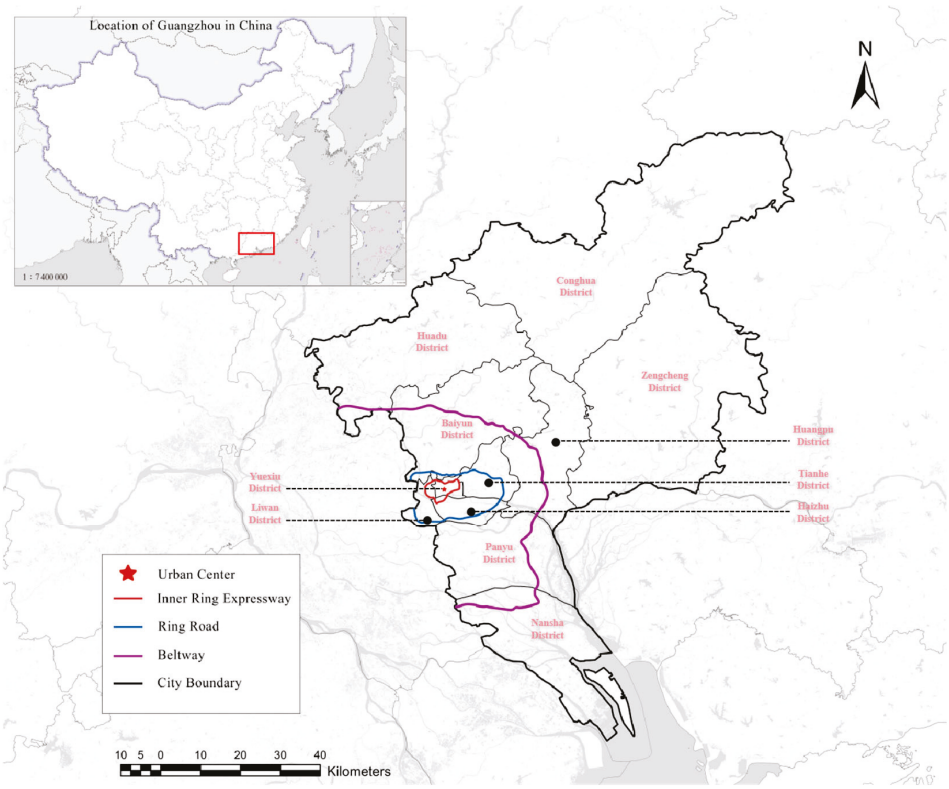


Figure 1. Location of Guangzhou, China.

The logistics industry in Guangzhou is undergoing a period of rapid growth, and the city currently leads the nation in terms of express-delivery business. In 2021, the express

business volume in Guangzhou reached 10.6 billion items, making it the first city in China to exceed this milestone [31]. This achievement underscores Guangzhou's pivotal role in the domestic express market. The city is home to the headquarters of several major express-delivery companies, including SF Express and YTO Express, which have extensive national operations. Additionally, Guangzhou features a well-developed express-delivery network and infrastructure, including numerous large logistics parks and distribution centers. Collectively, these assets provide a strong foundation for efficient and reliable express-delivery services.

In terms of industrial development, Guangzhou is renowned for its numerous "Taobao villages" and thriving wholesale markets. Taobao village refers to a village where the number of active online stores within the administrative village reaches more than 10% of the number of local households, or the annual transaction volume of e-commerce reaches more than 10 million yuan. These extensive commercial activities have significantly boosted the city's express-delivery industry. A recent study noted that Guangzhou was home to 273 Taobao villages in 2021 [31]; these conduct large-scale merchandising through e-commerce platforms, with Dayuan Village alone handling an average of over 3 million express parcels every day. Additionally, Guangzhou's wholesale industry is highly developed, as indicated by the presence of landmarks such as the Guangzhou International Commodity Exhibition Center and the Baiyun International Conference Center, as well as numerous professional markets offering a wide range of commodities to merchants nationwide. In 2020, the turnover of commodity trading markets in Guangzhou reached approximately 200 billion yuan, with individual market segments exceeding 100 million yuan. The city is highly regarded both domestically and internationally for its wholesale markets, which specialize in garments, electronics, and small commodities, attracting large numbers of buyers each year [32].

Guangzhou's unique attributes make it an ideal location for studying e-commerce demand. The city's vibrant economic environment, strategic geographic location, and advanced transportation and logistics infrastructure provide a compelling context for examining the spatial characteristics of parcel distribution.

3.2. Data Resources

3.2.1. Data Sources for E-Commerce Demand

In this study, e-commerce demand is operationalized as the spatial density of parcel deliveries received at the sub-district level, as recorded in express waybill data during the 2023 Singles' Day shopping festival. This metric serves as a proxy for aggregate online shopping activity, encompassing both individual consumer purchases and small-scale business transactions, but excluding bulk B2B logistics such as wholesale freight. Each parcel record corresponds to a single e-commerce transaction, regardless of the buyer type (consumer or micro-enterprise), thus reflecting total e-commerce-driven demand within a given spatial unit.

National express waybill data for this study were collected using the kuaidi100.com API to retrieve batch data, which was then processed using Python 3.11. The process entailed generating a batch of relevant waybill numbers based on express waybill-numbering regulations; the API was then called to retrieve tracking data using these numbers. Effective and authentic order numbers were filtered, and the obtained tracking data was then parsed to extract key information, including shipping locations, route points, receiving places, and timestamps.

The data sample of this study is focused on ZTO Express, which is a leading enterprise in the Chinese express delivery market with the highest market share (25%). As shown in Table 2, the example includes the waybill number, date, time, and process of the package

(where it was transported, in the state of collection, transfer, and receipt): The package was collected from Wanjiang, Guangzhou on 11 November 2023, and then transferred to the Guangzhou Transfer Center on the same day. It was then delivered to the Guangzhou Tonghe Express Point on 12 November 2023, and finally received at the Guangzhou Baiyun Dabi West Road Branch of Express Supermarket. The dataset included 3,342,444 records of ZTO Express waybill data from the Double 11 shopping event. Given that ZTO Express handled 145 million parcels on that day, this sample constitutes approximately 2.3% of the total volume, meaning it has the potential to provide statistically significant insights. For specific analysis of Guangzhou, 104,577 data points with Guangzhou as the receipt point were extracted. After excluding data with incomplete or unclear delivery details, 42,107 valid observations remained. This refined dataset was used for visualization and further analysis, as detailed in Table 2 and illustrated in Figure 2.

Table 2. Typical waybill data from ZTO Express.

Waybill Number	Date	Time	Progress
78XXXXXXXXXX28	11 November 2023	12:02:30	[Guangzhou Wanjia, Guangdong Province][Guangzhou] Guangzhou Wanjia has been collected
	11 November 2023	20:23:48	[Guangzhou Transfer Center, Guangdong Province][Guangzhou] The Guangzhou Express has arrived at the Guangzhou Transfer Center
	11 November 2023	20:29:38	[Guangzhou Transfer Center, Guangdong Province][Guangzhou] The package has been sent to Guangzhou Tonghe
	12 November 2023	1:04:23	[Guangzhou Tonghe, Guangdong Province][Guangzhou] The package has arrived at Tonghe, Guangzhou
	12 November 2023	10:12:25	[Guangzhou Tonghe, Guangdong Province][Guangzhou] The salesperson from Guangzhou Tonghe is currently delivering for the second time
	12 November 2023	15:37:04	[Guangzhou Baiyun Dabi West Road Branch of Express Supermarket, Guangdong Province] The package has been temporarily placed at the Guangzhou Baiyun Dabi West Road branch of the express supermarket, please pick up the package promptly.
	13 November 2023	14:11:58	[Guangzhou Baiyun Dabi West Road Branch of Express Supermarket, Guangdong Province][Guangzhou] Your package has been signed for

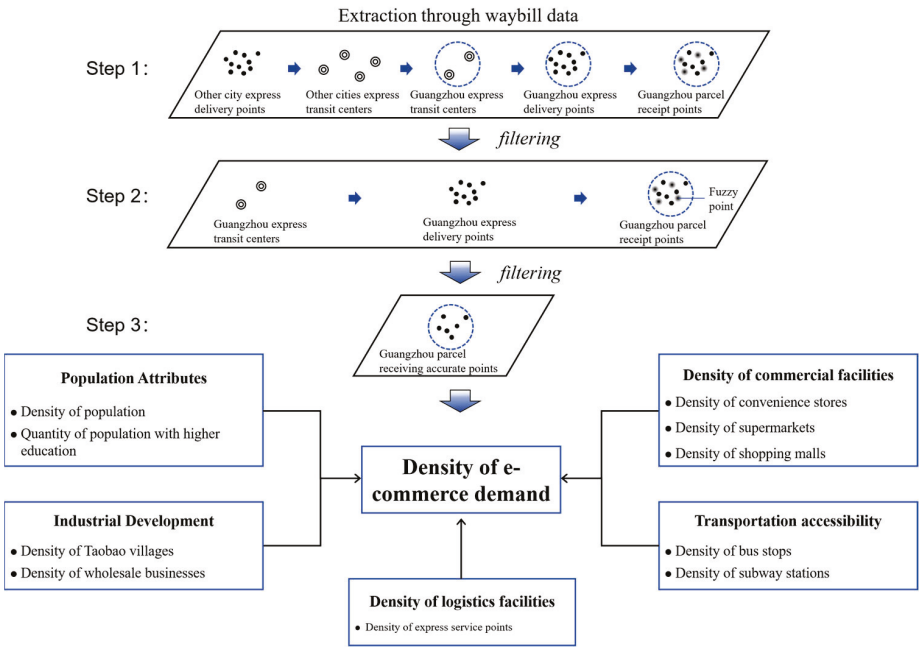


Figure 2. Illustration of conceptual framework.

3.2.2. Multi-Source Data Integration

The fundamental unit of analysis in this work was the sub-district or town, corresponding to the Chinese administrative terms "township" and "town." The selection of these observation units is critical because it directly impacts model outcomes. Larger units may fail to capture spatial heterogeneities, whereas very small units, despite their potential to enhance precision, can complicate data integration across diverse sources. By focusing on sub-districts or towns, which represent the smallest administrative units for socioeconomic indicators, this study achieves a balance between granular spatial analysis and practical data collection. Therefore, the analysis covered 176 sub-districts or towns in Guangzhou.

To gain a comprehensive understanding of the subject matter, a multi-source data collection approach was used in this study. Initially, detailed population-attribute data for each sub-district or township were obtained from publicly available records from the seventh population census of Guangzhou. This dataset included information on population size and education levels. Next, spatial-location data for Taobao villages in Guangzhou were acquired from the Ali Research Institute. Additionally, enterprise registration data were sourced from the industry and commerce department; this provided key details such as the names of enterprises, along with their business scope, registration dates, and addresses. The business-scope data highlighted the wholesale industry as a significant area of interest. Furthermore, geographical coordinates for all bus stops, subway stations, express service points, convenience stores, supermarkets, and shopping malls in Guangzhou were collected using point-of-interest data from the Amap open platform. To account for variations in the size of areas between streets, a density measure was employed to represent the distribution intensity of each variable within the research units. Descriptive statistics for the variables are summarized in Table 3.

Table 3. Descriptive statistics of variables.

Variable	Mean	Std. Dev.	Min	Max
Density of e-commerce demand (parcels/m^2)	0.415	0.455	0.0001	2.337
Density of population (persons/m^2)	174.062	191.065	0.614	975.420
Quantity of population with higher education (persons)	28,945.835	22,079.463	1034	119,755
Density of Taobao villages (villages/m^2)	0.0003	0.001	0	0.007
Density of wholesale businesses (points/m^2)	5.613	8.779	0.001	44.855
Density of bus stops (stops/m^2)	0.045	0.041	0	0.195
Density of subway stations (stations/m^2)	0.009	0.013	0	0.062
Density of express service points (points/m^2)	0.073	0.073	0.00004	0.458
Density of convenience stores (stores/m^2)	0.165	0.145	0	0.774
Density of supermarkets (stores/m^2)	0.030	0.031	0	0.221
Density of shopping malls (malls/m^2)	0.004	0.010	0	0.098

3.3. Methodology

This study utilizes waybill data to estimate e-commerce demand, facilitating a comprehensive comparison of global and local regression models. Parcel delivery volumes, operationalizing as a proxy for e-commerce demand, were analyzed as the dependent variable to capture spatially granular consumption behaviors. Explanatory variables were systematically categorized into factors related to population attributes, industrial development, the density of commercial facilities, transportation convenience, and the density

of logistics facilities. The detailed conceptual framework for these variables is shown in Figure 2.

3.3.1. Kernel Density Estimation

Kernel density estimation (KDE) is a widely used method for visualizing and analyzing spatial point distributions. They are applicable in various fields, including historical geography, crime analysis, public health, and transportation planning, among other topics within spatial humanities studies. Heatmaps use a gradient of warm and cool colors to illustrate the density distribution trends of point data, effectively minimizing overlap and highlighting areas of higher concentration. KDE is a classical spatial statistical method that generates heatmaps by applying a kernel function and attenuation effects to capture spatial point patterns, in accordance with Tobler's First Law of Geography [33]. In this study, KDE was employed to identify spatial heterogeneities in e-commerce demand. The general form of the KDE model is as follows:

$$f_n(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right) \tag{1}$$

where k is the weight function of the kernel; h is the bandwidth, i.e., the width of the surface extended in space with x as the origin, which affects the smoothness of the graph; and $x - x_i$ is the distance between the density-valuation point x and x_i.

3.3.2. Global Regression Modeling

In traditional regression approaches such as ordinary least square (OLS) regression, it is assumed that the relationship between independent and dependent variables remains consistent across the study area. In this study, population attributes, industrial development, the density of commercial facilities, transportation convenience, and the density of logistics facilities were selected as independent variables, with e-commerce demand as the dependent variable. The theoretical model delineating the relationships between the explanatory and dependent variables is as follows:

$$DPD_i = a_0 + b_1 \times PA_i + b_2 \times ID_i + b_3 \times CD_i + b_4 \times TC_i + b_5 \times PRD_i + \varepsilon \tag{2}$$

where i refers to the observation unit; DPD is the density of e-commerce demand; the population attributes PA include the population density and the quantity of population with higher education; the level of industrial development ID includes the density of Taobao villages and the density of wholesale businesses; the commercial-facilities density CD includes the density of convenience stores, supermarkets, and shopping malls; and the transportation convenience TC includes the density of bus stops and subway stations; PRD is the density of express service points; and ε is an estimation error term.

3.3.3. Local Regression Modeling

The global regression model is based on the assumption that the relationships between the explanatory and dependent variables are uniform and unchanging throughout the study area; this assumes that these relationships remain consistent across different spatial locations [34,35]. In contrast, a geographically weighted regression (GWR) model assumes that the regression coefficients will vary in different spatial locations. By incorporating spatial information, GWR models allow for more precise analysis of localized patterns that might be missed by global models. The structure of the GWR model used in this work is as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{m} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \tag{3}$$

where y_i refers to the explained variable, x_{ik} refers to the explanatory variables, (u_i, v_i) refers to the coordinates of the centroid of observation unit i, β_0 is an intercept term, and β_k is a regression coefficient for x_j.

Traditional GWR models are limited by their inability to regress all independent variables with varying bandwidths, and they can thus overlook local variations in features. In contrast, multiscale GWR (MGWR) models address this limitation by employing an optimal bandwidth for each independent variable. MGWR models allow for the selection of smaller bandwidths when independent variables exhibit significant local variation and larger bandwidths when they show more stable global patterns [36]. The equation for the MGWR model used in this work is

$$y_i = \sum_{j=0}^{k} \beta_{bwj}(u_i, v_i) x_{ij} + \varepsilon_i \tag{4}$$

where y_i is the explained variable, x_{ij} are the explanatory variables, (u_i, v_i) are to the coordinates of the centroid of observation unit i, bwj refers to the bandwidth used for the regression coefficient of x_j, and β_{bwj} is the regression coefficient for x_j.

4. Results

4.1. Kernel Density Estimation

The e-commerce demand in Guangzhou was found to exhibit pronounced spatial clustering, as shown in Figure 3. The highest e-commerce demand clustered in central Guangzhou and surrounding areas, reflecting the city's economic activity, population density, and transportation accessibility.

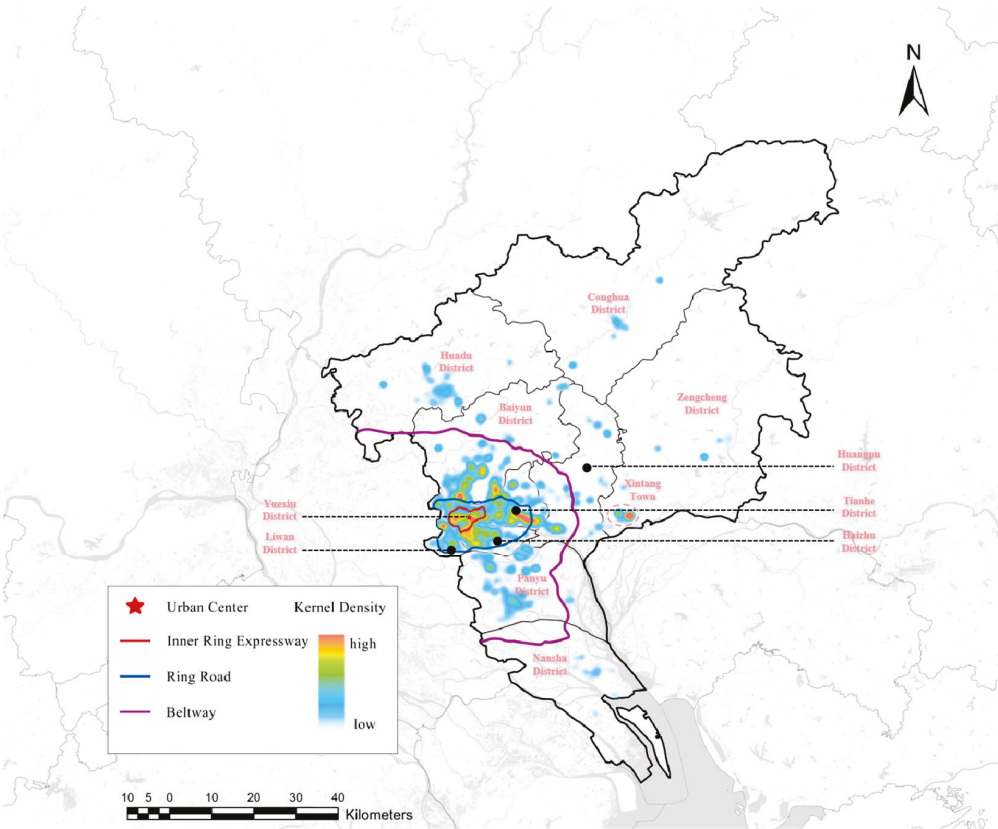


Figure 3. Spatial distribution of e-commerce demand obtained using KDE.

First, it is noted that the downtown area of Guangzhou, including Tianhe, Yuexiu, Liwan, Haizhu, and Baiyun districts, serves as the epicenter of economic activities. This concentration attracts significant commercial and residential demand, resulting in high-density e-commerce demand. The commercial prosperity and high population density in these districts are key drivers of the demand for parcels. Additionally, a noticeable concentration of high-value regions can be seen in the eastern part of Guangzhou, particularly in Zengcheng District, Xintang Town. This area is a critical node in the eastward "axis" of urban spatial development, and it acts as a gateway between Guangzhou and neighboring cities. Its strategic location and rapid growth contribute to a significant volume of parcel delivery.

Second, Guangzhou's well-developed transportation network, including the Inner Ring Expressway and main roads, facilitates the spatial aggregation of parcel delivery. The accessibility of these areas encourages logistics companies to set up facilities and distribution centers, thereby boosting the agglomeration effect of parcel intake.

The urban expansion pattern of Guangzhou has also influenced the spatial distribution of parcel delivery. The development of the city has transitioned from a concentric-circle model to a multi-core and belt-like pattern. This shift has resulted in the concentration of e-commerce demand not only in the city center but also in new agglomeration points between various core areas. The belt-shaped development along major transportation routes and rivers has further consolidated e-commerce demand due to the convenience of transportation links and the high concentration of commercial activities.

In summary, the spatial agglomeration of parcel delivery in Guangzhou is a result of the complex interplay between economic concentration, transportation accessibility, and evolving urban development patterns, exhibiting an overall trend of diffusion towards point-axis patterns. This confirms Anderson's theory of innovation diffusion [37], and on this basis, it can be concluded that the spatial pattern of e-commerce demand diffusion aligns with the spatial pattern of urban development.

4.2. Global Regression

To identify the factors associated with e-commerce demand, we included various socioeconomic and geographic variables in our modeling process and used stepwise regression to determine the most effective model. Variance inflation factor (VIF) analysis was used to assess multicollinearity among the independent variables, ensuring that only relevant predictors were included in the regression model.

As shown in Table 4, the VIF values for the explanatory variables ranged from 1.108 to 5.56, well below the multicollinearity threshold of 10 recommended by Salmerón et al. (2018) [38]. Following this, stepwise regression was conducted, and the model's fit was progressively improved with the addition of explanatory variables, as indicated by an increase in the adjusted R^2 value from 0.845 to 0.893. The analysis revealed that population attributes, industrial development, the density of logistics facilities, and the density of commercial facilities all significantly impact e-commerce demand. In contrast, transportation accessibility did not have a significant effect. Among the variables, population density, the proportion of the population with higher education, the density of Taobao villages, the density of express service points, and the density of shopping malls were found to have the most substantial influence on e-commerce demand.

Table 4. Summary statistics of the global OLS model.

Dimension	Variable	Model 1	Model 2	Model 3	Model 4	Model 5	VIF
Intercept		0.169 *	0.651	0.174	0.036	0.463	
Population attributes	Density of population	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.000 ***	3.313
	Quantity of population with higher education	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.000 ***	1.108
Industrial development	Density of Taobao villages		0.000 ***	0.000 ***	0.000 ***	0.000 ***	1.141
	Density of wholesale businesses		0.114	0.018 **	0.015 **	0.080 **	2.102
Transportation accessibility	Density of bus stops			0.622	0.888	0.336	5.052
	Density of subway stations			0.128	0.113	0.113	1.878
Density of logistics facilities	Density of express service points				0.073 *	0.002 ***	1.889
Density of commercial facilities	Density of convenience stores					0.285	5.115
	Density of supermarkets					0.031 **	5.560
	Density of shopping malls					0.000 ***	2.423
No. of Obs.		176	176	176	176	176	
Adjusted R^2		0.854	0.862	0.867	0.869	0.889	

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

It can be seen that population attributes significantly influence e-commerce demand. A higher population density indicates a larger number of residents in a given area, and this is naturally associated with increased consumption activities. Residents in densely populated areas are more likely to engage in frequent and diverse consumption, leading to a higher volume of parcel received. Additionally, a high population density is often correlated with advanced economic development and greater consumption capacity, further impacting e-commerce demand. Consistent with the conclusions drawn by other scholars, education level has a significant impact on e-commerce demand [21,39,40]. Individuals with higher education typically have higher incomes, which increases the likelihood of them purchasing more products and services, including online purchases. Having completed higher education is also correlated with familiarity and trust in the security and convenience of online shopping, which leads to more frequent online purchases and, consequently, a higher volume of parcels received. Moreover, highly educated individuals often hold knowledge-oriented or white-collar positions with demanding schedules, which may lead them to use parcel delivery services more frequently for home deliveries. Industrial development and logistics facilities also play a clear role in e-commerce demand.

Although it was initially hypothesized that transportation accessibility might reduce the frequency of people shopping online, thereby decreasing e-commerce demand due to increased offline shopping, the results indicate otherwise. While improved transportation options, such as bus and subway stations, may influence some individuals to shop offline more frequently, it was found that they do not directly affect the overall number of parcels received. Interestingly, the density of supermarkets and shopping malls significantly affects the volume of parcels received. It can be understood that supermarkets and shopping malls, as the main places for offline shopping, have advantages that are unmatched by online shopping, while they are usually located in populated and economically developed areas, which leads to a significant impact on e-commerce demand either positively or negatively.

The results of the global regression model reveal that Koenker's Breusch–Pagan statistic is statistically significant ($p < 0.01$), indicating regional disparities in the explanatory variables; this suggests the need for further local regression analysis to address these regional differences. As noted above, the OLS global model assumes a constant correlation between explanatory variables across the study area, but this fails to account for potential spatial variability in these relationships. To address this, Moran's I test was first conducted to examine spatial dependence in e-commerce demand. The Moran's I value was 0.551024 ($p < 0.001$), indicating a strong positive spatial autocorrelation in e-commerce demand. The findings underscore that the global OLS model, premised on spatial uniformity, fails to adequately capture the nuanced relationships between explanatory and dependent variables. Therefore, incorporating local regression models is essential to better account for spatial variations and improve the analysis.

4.3. Local Regression

To further investigate local spatial variation patterns in e-commerce demand, we incorporated the same explanatory variables from the global regression model into two local regression models: a GWR model and an MGWR model. Local regression models require a spatial weight matrix to account for spatial autocorrelation [41]. This adjacency-based weighting scheme aligns with spatial econometric principles [42], ensuring that neighborhood effects are captured without overextending spatial dependency assumptions at the municipal scale.

Table 5 presents a comparison of the fit between the local and global regression models. Overall, the local regression models show better fits than the global models. The adjusted R^2 value for the global regression model, which does not account for spatial correlation, is 0.889; the same value was obtained for the GWR model. In contrast, the MGWR model, which is optimized for bandwidth, achieved an adjusted R^2 value of 0.896. The OLS model, which is prone to overfitting, exhibits a negative corrected Akaike information criterion (AICc) value (−148) and a lower R^2 compared to both the GWR and MGWR models. The MGWR model, which has the highest adjusted R^2 (0.896) and a lower AICc value (126.923) than the GWR model, demonstrates a superior fit. This suggests that the MGWR model more effectively captures spatial heterogeneities than the GWR model.

Table 5. Comparison of the goodness-of-fit measures for the global and local models.

Criterion	OLS	GWR	MGWR
Adjusted R^2	0.889	0.889	0.896
AICc	−148.900	128.493	126.923

The bandwidths in the local regression models measure the influence ranges of independent variables on dependent variables. Table 6 displays the bandwidths for the variables in both the GWR and MGWR models. In the MGWR model, each independent variable has a unique bandwidth, while the GWR model uses a single fixed bandwidth. This allows the MGWR model to capture varying effective scales for different variables, whereas the GWR model only uses an average effective scale. The listed bandwidths indicate that the effective scales for the quantity of population with higher education and the density of supermarkets are relatively small, reflecting greater spatial variability in their impact on e-commerce demand. While the GWR model can only address spatial variation between variables, the MGWR model can offer insights into variation within variables. Consequently, the MGWR model offers a more nuanced analysis of spatial heterogeneities in the effects of independent variables.

Table 6. Multiscale bandwidth for the GWR and MGWR local models.

Criterion	Variable	GWR	MGWR
Intercept		138	175
Population attributes	Population density	138	175
	Quantity of population with higher education	138	93
	Quantity of population without higher education	138	175
Industrial development	Density of Taobao villages	138	175
	Density of wholesale businesses	138	175
Transportation accessibility	Density of bus stops	138	175
	Density of subway stations	138	175
Density of logistics facilities	Density of express service points	138	175
Density of commercial facilities	Density of convenience stores	138	175
	Density of supermarkets	138	112
	Density of shopping malls	138	175

As illustrated in Table 7, the influencing variables considered—the population density, the quantity of population with higher education, the density of express service points, the density of convenience stores, and the density of shopping malls—all exhibit positive correlations with e-commerce demand. The mean population density has the greatest influence (0.953) on the number of parcels received. The standard deviation (STD) of the quantity of population with higher education varies considerably, indicating that the impact of this variable differs notably across observation units. The density of Taobao villages, the density of wholesale businesses, and the density of supermarkets all have negative regression coefficients, indicating negative correlations with e-commerce demand. Based on this analysis, both the OLS and MGWR models lead to the same conclusion: the impact of transportation accessibility on e-commerce demand is not statistically significant.

Table 7. Summary of OLS and MGWR regression results.

Dimension	Variable	OLS	Mean	STD	Min	Median	Max
Intercept		0.463	0.058	0.002	0.052	0.059	0.062
Population attributes	Population density	0.000 ***	0.953	0.002	0.952	0.952	0.960
	Quantity of population with higher education	0.000 ***	0.136	0.073	0.014	0.141	0.265
Industrial development	Density of Taobao villages	0.000 ***	−0.067	0.001	−0.071	−0.067	−0.067
	Density of wholesale businesses	0.080 *	−0.092	0.002	−0.093	−0.093	−0.085
Transportation accessibility	Density of bus stops	0.366	−0.067	0.002	−0.069	−0.068	−0.061
	Density of subway stations	0.113	0.054	0.002	0.052	0.053	0.062
Density of logistics facilities	Density of express service points	0.002 ***	0.089	0.001	0.087	0.089	0.093
Density of commercial facilities	Density of convenience stores	0.285	0.120	0.004	0.116	0.119	0.137
	Density of supermarkets	0.031 **	−0.261	0.076	−0.341	−0.298	−0.128
	Density of shopping malls	0.000 ***	0.240	0.002	0.237	0.239	0.249
		OLS	Mean	STD	Min	Median	Max

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

This study makes a significant contribution to the literature by highlighting spatial heterogeneities in the influence of explanatory variables on e-commerce demand. Figure 4a shows that the coefficients of the MGWR model for population density vary spatially, with increasing differences from the city center to its periphery; however, the variation across different observation units remains relatively minor. This spatial heterogeneity can be attributed to the diverse lifestyles and shopping behaviors of residents in the city center, which diminish the impact of the population density in this area on e-commerce demand compared to in the periphery.

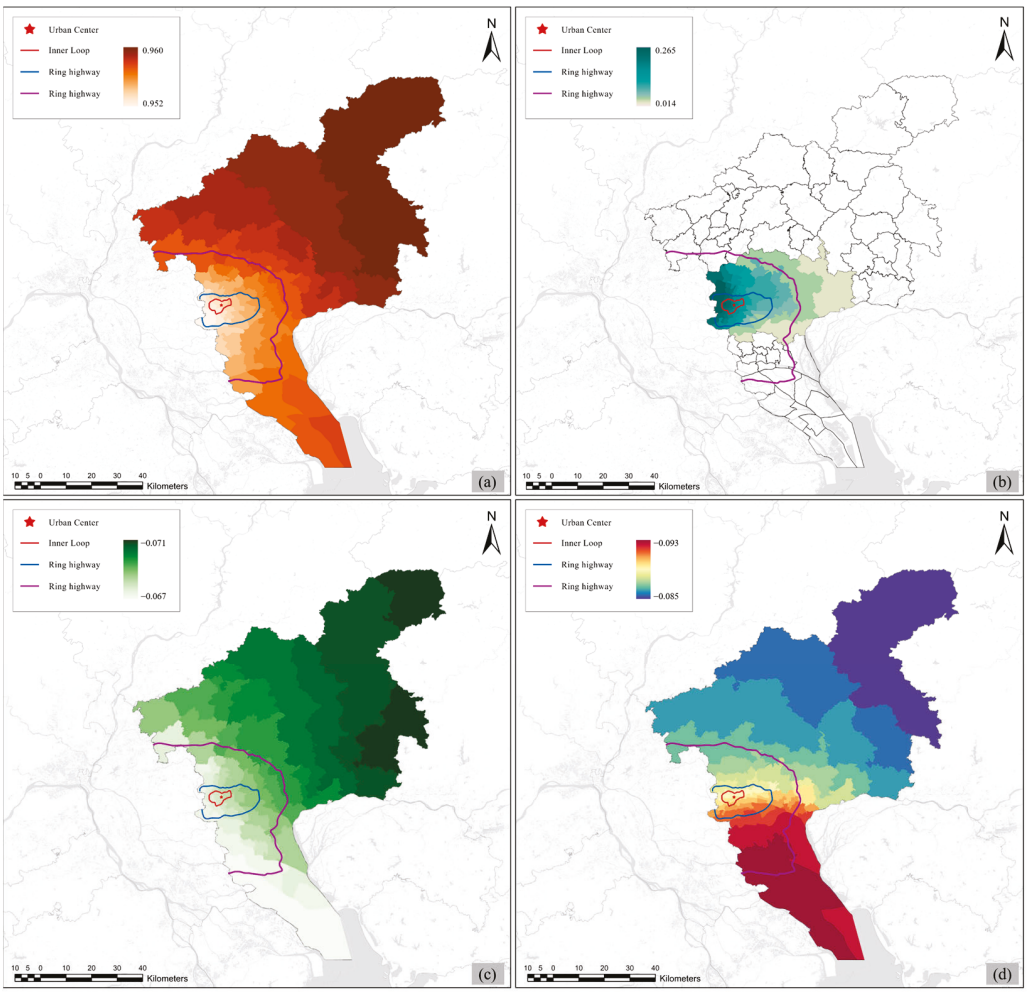


Figure 4. Spatial distributions of MGWR regression coefficients for population attributes and industrial development: (**a**) population density; (**b**) quantity of population with higher education; (**c**) density of Taobao villages; (**d**) density of wholesale businesses.

Figure 4b illustrates a decline in e-commerce demand from west to east after excluding insignificant regions, revealing more pronounced cellular variation. This variation can be linked to the prevalence of business activities and office areas in the western region. Additionally, the highly educated populations in these areas may have a greater tendency to use courier services for business-related items such as documents, samples, and contracts, leading to increased variability in parcel delivery.

Figure 4c reveals a negative correlation between the density of Taobao villages and e-commerce demand. This can be explained from two perspectives: first, Taobao's business model in Guangzhou ensures quality and minimizes returns, with a primary focus on dispatching rather than receiving parcels. Second, Taobao villages are typically located in

more remote areas where online shopping demand is low, resulting in a lesser impact on e-commerce demand.

Figure 4d shows a weak negative correlation between the density of wholesale businesses and e-commerce demand. This can be attributed to the nature of wholesale businesses, which generally handle commodities in bulk and undertake transactions that do not typically require courier or mail services. Instead, these transactions are often managed through specialized logistics channels such as freight and containerized transport. The negative correlation between density of wholesale businesses and e-commerce demand is more pronounced in the southern region compared to the northern region. In Guangzhou, the southern districts of Panyu and Nansha focus on imported goods, bulk items, and high-tech products, serving cross-border e-commerce and international trade. In contrast, the northern districts of Conghua and Huadu are known for the wholesale of flowers, daily necessities, and small commodities. This difference in commodity types and market positioning results in a weaker negative correlation between the density of wholesale businesses and parcel delivery in the northern part of Guangzhou.

Figure 5a illustrates that the impact of the density of express service points on e-commerce demand becomes more pronounced as one moves southward, although the effect remains relatively modest. In Guangzhou's southern districts, where logistics capacity is still nascent, even modest increases in the density of express service points yield disproportionately large gains in e-commerce parcel demand, underscoring its pronounced elasticity to supply. In contrast, despite having fewer express service points and less favorable economic conditions, the northern region benefits from the presence of convenience stores that partially meet the demand for parcel delivery. This supplementation reduces the overall impact of the density of express service points on e-commerce demand in the northern areas.

Figure 5b shows a more pronounced positive correlation between the density of convenience stores and e-commerce demand in the eastern, southern, and northern regions than in the western part of Guangzhou. In China, convenience stores often function as parcel-collection points; therefore, in areas with relatively few logistics facilities—such as the eastern, southern, and northern regions—the density of convenience stores has a stronger positive correlation with e-commerce demand. This effect is less evident in the western region, which has a higher density of logistics facilities. Additionally, supermarkets, which can provide alternatives to online shopping, contribute to this observed correlation. In general, the positive correlation between convenience store density and e-commerce demand underscores their role as critical last-mile touchpoints in omnichannel retail ecosystems.

After excluding the areas in which the results were not credible, Figure 5c reveals a negative correlation between supermarket density and e-commerce demand, with this correlation growing stronger from the urban core to the periphery. In peripheral urban areas, logistical challenges such as delayed deliveries and incomplete services can affect online shopping. Supermarkets in these suburban areas offer immediate access to goods—particularly for urgent needs or food items—leading customers to prefer in-person shopping over waiting for online deliveries. Thus, the density of supermarkets in these peripheral regions has a greater influence on e-commerce demand. Furthermore, the negative correlation between supermarket density and e-commerce demand provides empirical support for the viability of "click-and-collect" models [43], where in-store pickup substitutes for pure online delivery.

The results in Figure 5d align with the findings of the previous OLS model, showing a positive correlation between the density of shopping malls and e-commerce demand, with notable spatial heterogeneity from south to north. In the southern and central regions,

shopping malls are concentrated in commercial centers and feature large department stores and luxury brands, offering extensive offline shopping experiences. In response to the growth of online platforms, these malls are increasingly incorporating online shopping services and brand collaborations, thus increasing e-commerce demand. Conversely, due to geographic and demographic differences, shopping malls in the northern regions have adapted to offer more convenient and interactive shopping experiences; they also integrated online and offline marketing strategies earlier, promoting their products through e-commerce platforms. This integration has led to a spatial disparity between the north and south in terms of the influence of shopping malls on e-commerce demand.

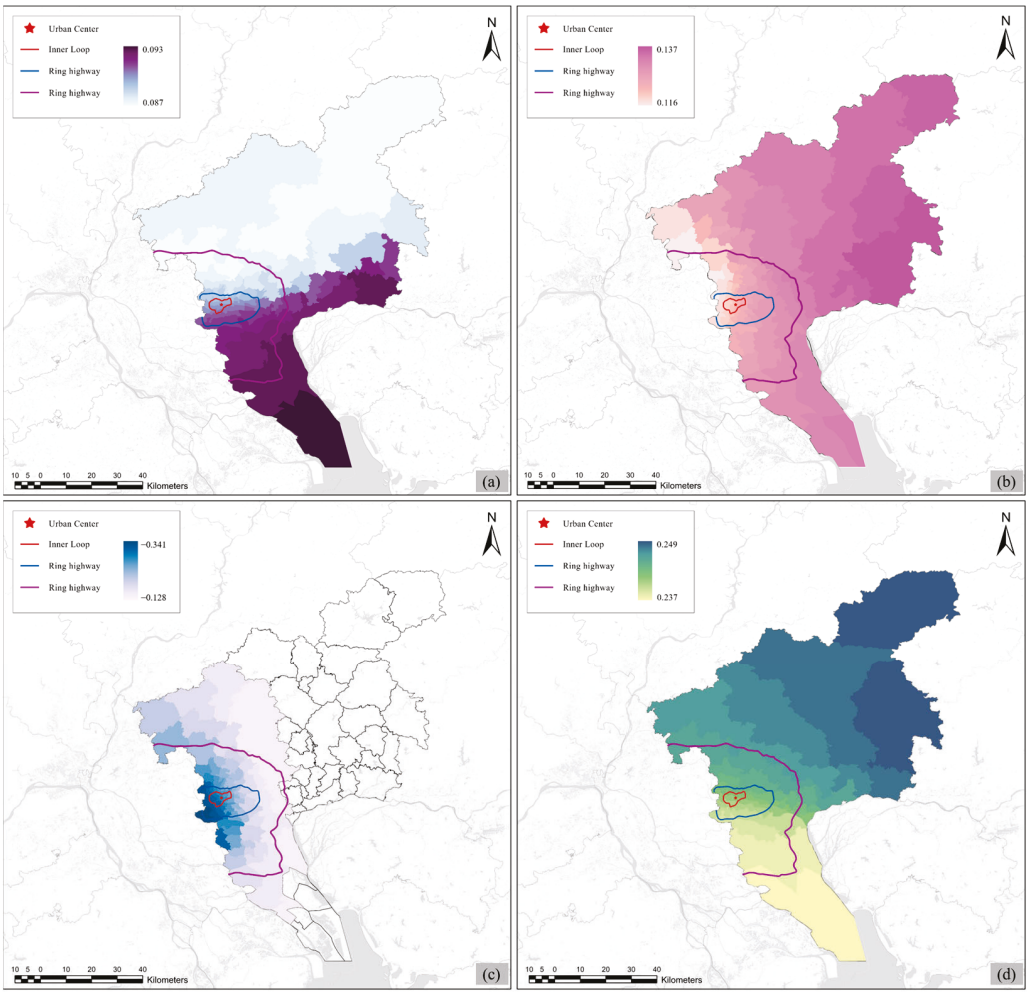


Figure 5. Spatial distributions of MGWR regression coefficients for logistics and the density of commercial facilities: (**a**) density of express service points; (**b**) density of convenience stores; (**c**) density of supermarkets; (**d**) density of shopping malls.

5. Conclusions and Discussion

This empirical study aims to characterize online shopping demand through e-commerce demand, explore the factors influencing its spatial distribution, and determine whether there is spatial heterogeneity among them. The data sources included waybill data for extracting parcel delivery activity and supplementary industry and geographic data. To test the hypothesis of the existence of spatial heterogeneities in the impacts of various factors on e-commerce demand, both global and local regression models were examined. While contextual factors (e.g., Taobao villages) are unique to China, our methodological framework and key findings—such as the dual role of convenience stores in logistics–retail integration and the substitutive effect of supermarkets—offer transferable insights for cities

in emerging economies with similar e-commerce growth trajectories (e.g., high-density Asian metropolises). Future comparative studies across cities could validate scalability.

Initially, global regression models were employed to analyze the relationships between e-commerce demand and several explanatory variables, including population attributes, industrial development, transportation accessibility, the density of logistics facilities, and the density of commercial facilities. This analysis revealed significant correlations between e-commerce demand and all of these variables except for transportation accessibility. This suggests that transportation accessibility does not directly affect online shopping habits or consumption behavior.

Subsequently, strong spatial autocorrelation was identified in e-commerce demand. Thus, the global model could not capture potential spatial correlations among explanatory variables, prompting the use of local regression models. The results indicate that local regression models offer a better fit compared to the original global model. Specifically, the MGWR model outperformed the GWR model, particularly regarding the effective bandwidth of the explanatory variables. The MGWR model confirmed substantial spatial variability in e-commerce demand. Among the factors analyzed, demographic attributes were found to have a more pronounced effect on e-commerce demand, showing significant spatial heterogeneities, while the density of logistical and commercial facilities provided additional insights.

Similar to previous research, there is a substitutive and complementary relationship between online shopping and physical stores [44–46]. At the same time, the necessity of considering retail geographic information methods in logistics planning was emphasized. Strategically co-locating convenience stores with parcel pickup services in peripheral areas can help address spatial inequities in service access while leveraging existing retail infrastructure. This approach can enhance last-mile delivery efficiency and improve consumer convenience, particularly in underserved regions. (1) The observed spatial heterogeneities indicate that the planning of logistics facilities should consider the influence of resource endowment, location conditions, economic factors, and other individual differences in different regions, and policies should be classified and implemented accordingly. Retailers and policymakers should consider spatially integrating retail facilities with logistics hubs to enhance consumer access to services while reducing delivery inefficiencies. For example, the co-location of supermarkets with parcel lockers or the expansion of convenience store pick-up networks can create synergistic service ecosystems while improving retail accessibility and delivery efficiency. (2) Our findings on the dual role of convenience stores align with the micro-fulfillment center concept [47], suggesting that these retail nodes not only serve immediate consumption needs but also act as decentralized logistics hubs, reducing last-mile delivery costs. China aims to achieve the goal of "countryside outlets and village services" by 2025. Strategic placement of convenience stores in peri-urban areas, informed by spatial demand elasticity patterns revealed in our MGWR analysis, could optimize last-mile network coverage while minimizing infrastructure redundancy. This approach will facilitate the efficient and compact establishment of small logistics service hubs. Such initiatives are expected to significantly enhance the supply chain resilience of economically disadvantaged and geographically isolated communities. (3) The negative correlation between the density of supermarkets and e-commerce demand shows the irreplaceable role of these shops, especially in people's daily lives when purchasing food and urgently needed products. Therefore, it is crucial to ensure comprehensive coverage of supermarkets when planning commercial facilities. These spatial strategies contribute to building retail service resilience by diversifying fulfillment channels, thereby mitigating supply chain disruptions during peak demand periods like shopping festivals. At the same time, this substitution effect mirrors the "showrooming vs. Webrooming" dynamic in omnichannel

retailing [48], where physical stores alternately compete with and complement digital platforms depending on product urgency and consumer time sensitivity.

This study's limitations stem from a sample based on a single promotional period that may not generalize to routine demand, the absence of a complete daily dataset, and the lack of SKU or user-level attributes; future work should (i) access the full daily dataset, (ii) apply the MGWR framework across multiple quarters or years, and (iii) merge anonymized mobile-app or survey data to capture product and consumer heterogeneity.

Author Contributions: Conceptualization, Y.C., J.C. and S.L.; methodology, J.C. and S.L.; software, J.C.; formal analysis, J.C. and S.L.; resources, Y.C. and S.L.; data curation, J.C.; writing—original draft preparation, J.C.; writing—review and editing, Y.C. and S.L.; visualization, J.C.; funding acquisition, Y.C. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China [Grant No. 42301211], National Key Research and Development Program of China [Grant No. 2022YFF1303105], and Science and Technology Program of Guangzhou (Grant No. 2025A04J5203).

Institutional Review Board Statement: Ethical review and approval were waived for this study, as it did not involve any interventions or procedures that required ethical approval.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jaller, M.; Pahwa, A. Evaluating the environmental impacts of online shopping: A behavioral and transportation approach. *Transport. Res. Part D-Transport. Environ.* **2020**, *80*, 15. [CrossRef]
2. SPB. Ti Insight Forecast Says Global Parcel Market Size Grows 9.2% This Year. Available online: https://www.spb.gov.cn/gjyzj/c200007/202408/c5d964def579416f8cf0cd97aaa3d0ba.shtml (accessed on 24 March 2025).
3. SPB. Global Express Development Report. Available online: http://www.spbdrc.org.cn/yzfzyjzx/c100010/202311/2dfe90054ff4494597b3d3de76448a1f.shtml (accessed on 24 March 2025).
4. Barthelmes, L.; Görgülü, M.E.; Kagerbauer, M.; Vortisch, P. Data collection for microscopic modelling of urban parcel transport to and from establishments—Empirical insights into city logistics in the region of Karlsruhe, Germany. *J. Urban Mobil.* **2024**, *6*, 8. [CrossRef]
5. Kim, S.-J.; Lim, H.; Park, M. Analysing the cost efficiency of parcel distribution networks with changes in demand. *Int. J. Urban Sci.* **2014**, *18*, 416–429. [CrossRef]
6. Lim, H.; Shiode, N. The impact of online shopping demand on physical distribution networks: A simulation approach. *Int. J. Phys. Distrib. Logist. Manag.* **2011**, *41*, 732–749. [CrossRef]
7. Aljohani, K.; Thompson, R.G. Impacts of logistics sprawl on the urban environment and logistics: Taxonomy and review of literature. *J. Transp. Geogr.* **2016**, *57*, 255–263. [CrossRef]
8. Veenstra, M.; Roodbergen, K.J.; Coelho, L.C.; Zhu, S.X. A simultaneous facility location and vehicle routing problem arising in health care logistics in the Netherlands. *Eur. J. Oper. Res.* **2018**, *268*, 703–715. [CrossRef]
9. Zhuang, Y.F.; Huang, R. Design and Implementation of Terminal Logistics Distribution System Based on IOS Platform. In Proceedings of the International Conference on Electronic Information Technology and Computer Engineering (EITCE), Beijing Normal Univ, Zhuhai, China, 23–24 September 2017.
10. Seghezzi, A.; Siragusa, C.; Mangiaracina, R. Parcel lockers vs. home delivery: A model to compare last-mile delivery cost in urban and rural areas. *Int. J. Phys. Distrib. Logist. Manag.* **2022**, *52*, 213–237. [CrossRef]
11. Kim, J.; Kim, M.; Choi, J.; Trivedi, M. Offline social interactions and online shopping demand: Does the degree of social interactions matter? *J. Bus. Res.* **2019**, *99*, 373–381. [CrossRef]
12. Reiffer, A.S.; Kuebler, J.; Kagerbauer, M.; Vortisch, P. Agent-based model of last-mile parcel deliveries and travel demand incorporating online shopping behavior. *Res. Transp. Econ.* **2023**, *102*, 101368. [CrossRef]
13. Wang, J.J.; Xiao, Z.P. Co-evolution between etailing and parcel express industry and its geographical imprints: The case of China. *J. Transp. Geogr.* **2015**, *46*, 20–34. [CrossRef]
14. Hou, J.W.; Elliott, K. Mobile shopping intensity: Consumer demographics and motivations. *J. Retail. Consum. Serv.* **2021**, *63*, 11. [CrossRef]

15. Cheng, C.; Sakai, T.; Alho, A.; Cheah, L.; Ben-Akiva, M. Exploring the Relationship between Locational and Household Characteristics and E-Commerce Home Delivery Demand. *Logistics* **2021**, *5*, 29. [CrossRef]
16. de Sousa, L.T.M.; de Oliveira, L.K.; dos Santos, J.L., Jr.; Bertoncini, B.V.; Isler, C.A.; Larranaga, A.M. Influence of neighborhood characteristics on e-commerce deliveries: The case of Belo Horizonte, Brazil. *Res. Transp. Econ.* **2023**, *100*, 12. [CrossRef]
17. Oliveira, L.K.; Sousa, L.T.M.; Santos, J.L., Jr.; Bracarense, L. Spatial fragment parcel deliveries landscape in a global south city: The case of Belo Horizonte, Brazil. *Transp. Res. Part A-Policy Pract.* **2025**, *199*, 104571. [CrossRef]
18. Shin, N.; Lim, H.; Kim, Y.J. Modeling spatial dimensions of parcel delivery demand and its determinants. *Int. J. Logist.-Res. Appl.* **2023**, *26*, 17. [CrossRef]
19. Shao, R.; Derudder, B.; Witlox, F. The geography of e-shopping in China: On the role of physical and virtual accessibility. *J. Retail. Consum. Serv.* **2022**, *64*, 10. [CrossRef]
20. Farag, S.; Weltevreden, J.; van Rietbergen, T.; Dijst, M.T.; van Oort, F. E-shopping in the Netherlands: Does geography matter? *Environ. Plan. B-Plan. Des.* **2006**, *33*, 59–74. [CrossRef]
21. Rotem-Mindali, O. E-tail versus retail: The effects on shopping related travel empirical evidence from Israel. *Transp. Policy* **2010**, *17*, 312–322. [CrossRef]
22. Beckers, J.; Cardenas, I.; Verhetsel, A. Identifying the geography of online shopping adoption in Belgium. *J. Retail. Consum. Serv.* **2018**, *45*, 33–41. [CrossRef]
23. Petcharat, T.; Jattamart, A.; Leelasantitham, A. A conceptual model to imply a negative innovation assessment framework on consumer behaviors through the electronic business platforms. *J. Retail. Consum. Serv.* **2023**, *74*, 20. [CrossRef]
24. Davari, A.; Iyer, P.; Rokonuzzaman, M. Identifying the determinants of online retail patronage: A perceived risk perspective. *J. Retail. Consum. Serv.* **2016**, *33*, 186–193. [CrossRef]
25. Cao, X.Y.; Chen, Q.; Choo, S.H. Geographic Distribution of E-Shopping Application of Structural Equation Models in the Twin Cities of Minnesota. *Transp. Res. Rec.* **2013**, *2383*, 18–26. [CrossRef]
26. Clarke, G.; Christopher, T.; Birkin, M. The emerging geography of e-commerce in British retailing. *Reg. Stud. Reg. Sci.* **2015**, *2*, 371–391. [CrossRef]
27. Ren, F.; Kwan, M.P. The impact of geographic context on e-shopping behavior. *Environ. Plan. B-Plan. Des.* **2009**, *36*, 262–278. [CrossRef]
28. Loo, B.P.Y.; Wang, B. Factors associated with home-based e-working and e-shopping in Nanjing, China. *Transportation* **2018**, *45*, 365–384. [CrossRef]
29. Etminani-Ghasrodashti, R.; Hamidi, S. Online shopping as a substitute or complement to in-store shopping trips in Iran? *Cities* **2020**, *103*, 16. [CrossRef]
30. GSB. Bulletin of the Seventh National Population Census of Guangzhou Municipality. Available online: https://tjj.gz.gov.cn/zzfwzq/tjgb/content/post_8543052.html (accessed on 24 March 2025).
31. GMPA. Guangzhou Postal Industry Development Statistics Bulletin 2021. Available online: http://gd.spb.gov.cn/gdsyzglj/c106014/c106030/c106036/202205/79ba2af477454cbabc795453d6c4e56d.shtml (accessed on 24 March 2025).
32. GSB. Market Advantages of Good Toughness and Resilience—Analysis of the Current Situation, Problems and Countermeasures of Guangzhou Billion Dollar Market Development in the "13th Five-Year Plan" Period. Available online: https://tjj.gz.gov.cn/zzfwzq/sjjd/content/post_8763121.html (accessed on 24 March 2025).
33. Yuan, K.; Cheng, X.; Gui, Z.; Li, F.; Wu, H. A quad-tree-based fast and adaptive Kernel Density Estimation algorithm for heat-map generation. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 2455–2476. [CrossRef]
34. Brunsdon, C.; Fotheringham, A.S.; Charlton, M.E. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geogr. Anal.* **1996**, *28*, 281–298. [CrossRef]
35. Fotheringham, A.S.; Charlton, M.E.; Brunsdon, C. Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis. *Environ. Plan. A* **1998**, *30*, 1905–1927. [CrossRef]
36. Fotheringham, A.S.; Yang, W.; Kang, W. Multiscale Geographically Weighted Regression (MGWR). *Ann. Am. Assoc. Geogr.* **2017**, *107*, 1247–1265. [CrossRef]
37. Anderson, W.P.; Chatterjee, L.; Lakshmanan, T.R. E-commerce, Transportation, and Economic Geography. *Growth Change* **2003**, *34*, 415–432. [CrossRef]
38. Salmerón, R.; García, C.B.; García, J. Variance Inflation Factor and Condition Number in multiple linear regression. *J. Stat. Comput. Simul.* **2018**, *88*, 2365–2384. [CrossRef]
39. Doan, D.A.; Vu, N.H.; Nguyen, P.L.; Dinh, D.X. Vietnamese people's behavior and future intention to purchase medicines and functional foods on the internet: A cross-sectional study. *Sci Rep* **2024**, *14*, 10. [CrossRef]
40. Roos, J.M.; Jansson, M.; Bjerkeli, P.J. Who Are the Online Medication Shoppers? A Market Segmentation of the Swedish Welfare State. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 526–537. [CrossRef]
41. Brunsdon, C.; Fotheringham, A.S.; Charlton, M. Geographically weighted summary statistics—A framework for localised exploratory data analysis. *Comput. Environ. Urban Syst.* **2002**, *26*, 501–524. [CrossRef]

42. Tobler, W.R. A Computer Movie Simulating Urban Growth in the Detroit Region. *Econ. Geogr.* **1970**, *46*, 234–240. [CrossRef]
43. Milioti, C.; Pramatari, K.; Kelepouri, I. Modelling consumers' acceptance for the click and collect service. *J. Retail. Consum. Serv.* **2020**, *56*, 8. [CrossRef]
44. Ding, Y.; Lu, H.P. The interactions between online shopping and personal activity travel behavior: An analysis with a GPS-based activity travel diary. *Transportation* **2017**, *44*, 311–324. [CrossRef]
45. Lee, R.J.; Sener, I.N.; Mokhtarian, P.L.; Handy, S.L. Relationships between the online and in-store shopping frequency of Davis, California residents. *Transp. Res. Part A-Policy Pract.* **2017**, *100*, 40–52. [CrossRef]
46. Wang, K.L.; Gao, Y.; Liu, Y.C.; Habib, K.N. Exploring the choice between in-store versus online grocery shopping through an application of Semi-Compensatory Independent Availability Logit (SCIAL) model with latent variables. *J. Retail. Consum. Serv.* **2023**, *71*, 17. [CrossRef]
47. Kim, J.; Choi, J.; Chang, S.R.; Trivedi, M. The impact of offline store presence on digital sales: The moderating role of product functionality. *J. Retail. Consum. Serv.* **2024**, *78*, 10. [CrossRef]
48. Flavián, C.; Gurrea, R.; Orús, C. Combining channels to make smart purchases: The role of webrooming and showrooming. *J. Retail. Consum. Serv.* **2020**, *52*, 11. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

systems

MDPI

Article

Identification and Prediction Methods for Frontier Interdisciplinary Fields Integrating Large Language Models

Yu Wu [1], Qiao Lin [1], Jinming Wu [1], Ru Yao [2] and Xuefu Zhang [1,*]

[1] Agricultural Information Institute of Chinese Academy of Agricultural Sciences, Beijing 100081, China; wuyu20@mails.ucas.ac.cn (Y.W.); linqiao@caas.cn (Q.L.); wujinming@caas.cn (J.W.)
[2] Institute of Data Science and Agricultural Economy, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; yaoru@baafs.net.cn
* Correspondence: zhangxuefu@caas.cn

Abstract: Identifying frontier interdisciplinary domains is essential for tracking scientific evolution and informing strategic research planning. This study proposes a comprehensive framework that integrates (1) semantic disciplinary classification using a large language model (GPT-3.5-Turbo), (2) quantitative metrics for interdisciplinarity (degree and integration strength) and frontierness (novelty, growth, and impact), and (3) trend prediction using time series models, including Transformer, LSTM, GRU, Random Forest, and Linear Regression. The framework systematically captures both structural and temporal dimensions of emerging research fields. Compared to conventional citation-based or topic modeling approaches, it enhances semantic precision, supports multi-label classification, and enables forward-looking forecasts. Empirical validation shows that the Transformer model achieved the highest predictive performance, outperforming other deep learning and baseline models. As an illustrative example, the framework was applied to synthetic biology, which demonstrated high interdisciplinarity, strong novelty, and growing academic influence. These results underscore the field's strategic position as a frontier interdisciplinary domain. Beyond this case, the proposed framework is generalizable to other domains and provides a scalable, data-driven solution for dynamic monitoring of emerging interdisciplinary areas. It holds promise for applications in science and technology intelligence, research evaluation, and policy support.

Keywords: frontier research; interdisciplinarity; deep learning forecasting; transformer model; trend prediction; science and technology intelligence

1. Introduction

With the rapid development of science and technology, the boundaries between disciplines have gradually become blurred [1], leading to an increasing trend toward interdisciplinary and integrative scientific research. Interdisciplinary research has emerged as a critical domain for innovation breakthroughs [2], especially in addressing complex global challenges such as climate change [3], biotechnology [4], and artificial intelligence [5]. In this context, interdisciplinary approaches have created significant opportunities for innovation. At the same time, frontier scientific fields play an essential role in driving technological advancement [6] and tackling global issues. As emerging fields continue to evolve, the timely identification of potential frontier interdisciplinary areas can provide valuable guidance for policymakers, scholars, and industries, enabling the strategic allocation of research resources in advance [7,8].

Frontier interdisciplinary fields are characterized by both "frontierness" and "interdisciplinarity". They represent the cutting edge of scientific exploration, integrating the latest achievements and technologies from different disciplines, supporting breakthrough innovation, and promoting the birth of new knowledge. These fields significantly expand the boundaries of scientific research and deepen its applications [9,10]. However, effectively identifying these potential frontier interdisciplinary areas from the vast amount of academic output and predicting their future development trajectories remains a major challenge in current research [11,12].

At present, methods for identifying frontier fields largely rely on citation analysis [13], textual content mining [14], and expert judgment [15]. Although these approaches have, to some extent, revealed patterns of field evolution, they also face pressing limitations. For instance, citation analysis suffers from issues such as self-citation anomalies [16] and time-lag effects [17]. Content analysis, often based on topic modeling, is known to encounter serious conceptual and practical problems, including the lack of well-grounded Bayesian priors, discrepancies with the true statistical characteristics of text data, and difficulties in accurately determining the number of topics [18]. Meanwhile, expert judgment is inherently subject to biases and subjectivity [19].

Current approaches to identifying frontier interdisciplinary fields primarily rely on topic modeling [14,20], focusing on subfield-level analysis while lacking mechanisms to evaluate and forecast the overall interdisciplinarity and frontierness of entire research domains.

To address this gap, we propose an integrated framework that combines large language models, quantitative metrics, and time series forecasting. GPT-3.5-Turbo is used to classify multidisciplinary papers from the Web of Science, enabling fine-grained disciplinary tagging. Interdisciplinarity is measured through both static indicators—such as the degree and strength of cross-disciplinary integration—and dynamic network-based metrics. Frontierness is assessed using a composite index of novelty, growth, and impact. Time series models are then compared to select the optimal method for forecasting development trajectories.

Applying this framework to synthetic biology, we demonstrate its utility in identifying frontier interdisciplinary domains, supporting evidence-based decisions in research management and innovation policy. This study offers not only a new methodological pathway for frontier identification but also practical tools for guiding interdisciplinary innovation and optimizing resource allocation.

The proposed framework offers both a novel technological approach for quantitative frontier identification and critical decision support for research management, resource allocation, and the construction of innovation systems. It holds significant implications for advancing academic innovation and driving societal progress.

This study develops a structured and integrative framework that consolidates existing methods—LLM-based semantic classification, multidimensional frontierness evaluation, and time series trend forecasting—into a coherent pipeline aimed at supporting the strategic identification of frontier interdisciplinary domains. The value of this framework lies in its ability to systematically combine and apply recent advances across several methodological strands. Specifically, the framework offers the following:

(1) Methodological Innovation: We propose an integrated framework that combines semantic classification via ChatGPT-3.5-Turbo, novel interdisciplinarity and frontierness metrics, and time series forecasting. This overcomes the limitations of traditional citation-based or keyword-based approaches and enhances the precision and scalability of frontier field identification.
(2) Metric System for Interdisciplinarity and Frontierness: A set of indicators is introduced to quantify the degree and depth of disciplinary integration, as well as the

novelty, growth, and impact of a field. These metrics provide a systematic basis for identifying and comparing interdisciplinary dynamics across domains.

(3) Time Series Forecasting for Frontier Trend Analysis: We evaluate and compare linear, machine learning, and deep learning models for predicting research field trajectories. The Transformer model is found to perform best, offering a robust, data-driven tool for prospective trend analysis in science and technology studies.

(4) Empirical Validation: Using synthetic biology as a case study, the framework demonstrates its practical effectiveness in capturing the interdisciplinary structure, research frontierness, and developmental trajectory of an emerging field.

This study contributes both methodologically and practically to the field of science and technology intelligence, offering a decision-support tool for researchers, institutions, and policymakers in managing and anticipating interdisciplinary innovation.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the proposed methodology. Section 4 reports the empirical results. Section 5 discusses the methodology and research findings. Finally, Section 6 concludes the study.

2. Literature Review

2.1. Research Front Identification

Frontier research generally refers to the most innovative and promising directions within a discipline or across interdisciplinary fields [7]. Accurately identifying research frontiers is crucial for research management agencies and policymakers, as it facilitates the optimization of resource allocation and drives breakthroughs in key technologies [21,22]. For researchers, understanding and grasping research frontiers not only ensures the academic value and practical impact of their work but also promotes interdisciplinary collaboration and knowledge innovation [23].

In bibliometric studies, "Research Fronts" is a key concept for assessing and identifying frontier areas. Price de Solla [24] was among the first to propose that focusing on approximately 30 to 50 recently published highly cited papers and their associated research topics within a domain could effectively reflect the research frontiers of that field. Later, Small and Griffith [25] further defined research fronts as clusters of highly cited or co-cited documents.

Subsequently, scholars have proposed a variety of methods for frontier identification, mainly categorized into qualitative methods based on expert knowledge and quantitative methods leveraging computational technologies [26]. Among qualitative approaches, the Delphi method is a classic and widely applied technique for identifying frontier interdisciplinary fields [27]. Quantitative approaches, which have become a research focus, include citation network-based analysis and content-based analysis [26].

Citation analysis mainly includes co-citation analysis [28] and bibliographic coupling [29]. Small and Griffith [25] Small and Griffith (1974) proposed that co-cited documents often share certain content similarities. Through co-citation analysis, the intrinsic relationships between documents can be revealed, and by applying clustering methods [30,31] and community detection techniques [32], researchers can further explore research fronts within academic fields.

On the other hand, bibliographic coupling [33] posits that documents sharing common references are likely to be closely related in content. Based on this concept, numerous studies have explored frontier topics through bibliographic coupling [33,34]. For instance, Liu et al. [35] used bibliographic coupling networks to detect scientific research fronts and trends, while Wei et al. [36] applied this approach to investigate research fronts in low-carbon technologies. Although these methods have achieved certain successes in revealing the dynamics of disciplinary development, they commonly suffer from time lag

issues: newly published papers need time to accumulate citations, making it difficult to promptly capture emerging research areas and potential frontiers.

Content-based approaches, in contrast, directly analyze the textual content of documents to identify research fronts. Content analysis typically operates at two granularities: lexical and thematic. From the lexical perspective, researchers detect emerging terms and concepts in papers, patents, and other texts, or apply co-word analysis methods. For example, Kleinberg [37] proposed a burst detection algorithm capable of quickly capturing the emergence of research hotspots or new ideas in a field. Li and Chu [38] explored the identification of specific research frontiers based on enhanced co-word analysis techniques. Li and Chu [38] explored the identification of specific research frontiers based on enhanced co-word analysis techniques. From the thematic perspective, studies have employed models such as Latent Dirichlet Allocation (LDA) [39–41] and BERTopic [42] for thematic frontier discovery. Within the framework of topic modeling, several studies have further proposed corresponding frontier evaluation indicators encompassing dimensions such as impact, novelty, and growth [43,44].

2.2. Interdisciplinary Research Identification

Given the close methodological association between frontier detection and interdisciplinarity assessment, certain techniques such as topic modeling and citation analysis are discussed in both sections, though applied to different analytical purposes.

Research on interdisciplinary identification has increasingly adopted topic modeling approaches, either by first identifying interdisciplinary documents or by directly evaluating topic-level interdisciplinarity [45]. These methods often draw from keyword co-occurrence, semantic modeling, and citation-based linkages—many of which are also used in research front detection.

Keyword mining uses extracted terms to infer interdisciplinary trends. For instance, Xu et al. [46] and Wang et al. [47] analyzed keyword diversity and introduced measures such as "topic term interdisciplinarity" [48]. While intuitive and scalable, this approach is sensitive to keyword quality and struggles with synonym recognition and latent semantic links.

Topic modeling, especially using LDA or BERTopic, has become a mainstream method for capturing thematic interdisciplinarity. Recent work has integrated semantic embeddings [49], cosine similarity and Rao–Stirling diversity [50], and even graph neural networks [51] to enhance topic coherence and interdisciplinary linkage prediction. However, these models often face challenges in dynamic tracking and interpretability, particularly when modeling the evolving nature of interdisciplinary fields.

In the area of citation network analysis, citation-based interdisciplinary topic identification methods are categorized into co-citation analysis, bibliographic coupling, and direct citation analysis [45]. Omodei et al. [52] proposed a bipartite multilayer network analysis approach based on citations and disciplines to assess the interdisciplinary importance of scholars, institutions, and countries. Adams and Light [53] constructed bibliographic coupling networks for AIDS research papers and identified topics spanning multiple disciplinary communities as interdisciplinary topics.

Currently, interdisciplinary field identification methods exhibit a trend of increasing diversification. Nevertheless, each category of methods faces inherent limitations. Keyword mining depends on the precision and completeness of terms and struggles to detect implicit interdisciplinary linkages. Text content mining approaches (e.g., BERTopic, graph neural networks) offer advantages in semantic understanding and knowledge recombination but still face challenges regarding model interpretability and dynamic tracking. Citation analysis methods effectively reveal knowledge flows and disciplinary associations;

however, due to the time lag inherent in citation accumulation, they are less capable of reflecting the latest frontier dynamics.

2.3. Time Series Modeling in Interdisciplinary Frontier Identification

Time series analysis, as a predictive method, has been widely applied in scientific trend analysis in recent years [54–56]. Particularly in the prediction of disciplinary frontiers, it enables the identification of development trajectories across different fields [57], thereby facilitating the targeted allocation of scientific and technological resources [58]. The basic idea of time series analysis is to use historical data to build models that infer future trends, making it particularly suitable for complex systems characterized by continuity and cumulative effects.

2.4. Limitations of Existing Research

Despite notable advancements, current studies on the identification and prediction of frontier fields still face several limitations:

(1) Methodological Fragmentation:

Most approaches rely on a single method—such as bibliometrics, machine learning, or topic modeling—without integrating their respective strengths. This constrains the capacity to conduct comprehensive, multidimensional analyses of interdisciplinarity and frontier dynamics.

(2) Narrow Scope of Interdisciplinary Focus:

Existing research tends to emphasize subfield-level topic detection, lacking tools for identifying and modeling interdisciplinarity at the domain level. As disciplinary boundaries become increasingly fluid, there is a growing need for more holistic and scalable frameworks.

(3) Static Analysis and Forecasting Gaps:

Many studies are based on retrospective data, with limited adoption of predictive models. The absence of dynamic forecasting restricts their utility for prospective technology assessment and policy planning.

(4) Semantic and Temporal Limitations:

Conventional methods (e.g., keyword mining, LDA, citation analysis) often struggle with semantic nuance, time sensitivity, and early detection of emerging topics. They are hindered by citation delays, rigid topic structures, and limited ability to capture deep contextual interdisciplinarity.

To clearly present the strengths and weaknesses of existing methods, we summarize them in the Table 1 below:

Table 1. Comparison of common approaches for frontier field identification.

Method	Advantages	Limitations
Citation Analysis	Reveals knowledge flow; quantifiable; useful for structural mapping	Time lag in citation accumulation; sensitive to self-citations
Keyword/Content Mining	Captures semantic patterns; scalable; identifies emerging terms	Dependent on keyword quality; limited semantic depth; difficulty with synonyms
Topic Modeling (e.g., LDA, BERTopic)	Detects thematic structure; supports trend analysis	Struggles with dynamic evolution; interpretability challenges; topic count sensitivity
Expert Judgment	Leverages domain knowledge; useful for early-stage fields	Subjective bias; lacks scalability; low reproducibility

In response, this study proposes an integrated framework combining large language models, bibliometric metrics, and time series forecasting to enable more accurate and dynamic identification of frontier interdisciplinary domains.

3. Research Framework

Identifying frontier interdisciplinary domains requires integrating three core dimensions: interdisciplinarity, frontierness, and development trends. Interdisciplinarity reflects cross-domain integration, while frontierness captures innovation potential and emerging influence. By combining LLM-based semantic classification, bibliometric analysis, and trend forecasting, this study proposes a unified framework to systematically identify high-potential domains and support strategic decision-making in science and technology development.

The research process is divided into five major steps (Figure 1): (1) Data Collection and Disciplinary Classification: Extract paper data from the WOS database covering the past ten years to construct a broad academic dataset; utilize OpenAI's GPT-3.5-Turbo for disciplinary classification. (2) Interdisciplinarity Identification: Based on the classification results, design interdisciplinarity measurement indicators to calculate the degree of interdisciplinarity and the strength of interdisciplinary integration. (3) Frontierness Assessment: Evaluate the frontierness of each domain by integrating three dimensions: influence, growth, and novelty. (4) Field Trend Analysis: Employ five different time series models to analyze the future trends in the academic impact of papers from each field, modeling and forecasting future development trajectories. (5) Identification of Frontier Interdisciplinary Fields: Integrate interdisciplinarity and frontierness indicators, combined with the technological trend analysis, to identify potential frontier interdisciplinary fields, thereby providing empirical support for technological innovation and policymaking within the respective fields.

3.1. Interdisciplinarity Identification Method

This study analyzes interdisciplinarity at three levels: (1) disciplinary classification, (2) static indicator construction, and (3) dynamic network analysis. First, we use ChatGPT to perform semantic classification of papers, allowing for the identification of multiple relevant disciplines and addressing the limitations of traditional single-label systems. Second, we construct interdisciplinarity metrics based on bibliometric distributions to quantify integration breadth and depth. Third, we extract dynamic indicators from both static and temporal co-occurrence networks to reveal the structural evolution of interdisciplinary relationships over time.

3.1.1. Paper Classification System Based on ChatGPT

Conventional classification systems such as Web of Science (WoS), Scopus, and Fields of Research (FOR) offer standardized, journal-level taxonomies with broad coverage; however, they face limitations in handling interdisciplinary and heterogeneous data. These systems often assign a single discipline per document, lack contextual understanding, and require manual effort, making them less effective for large-scale or interdisciplinary analyses.

To address these issues, this study employs GPT-3.5-turbo-0125 for paper classification. As a large language model, ChatGPT enables deep semantic understanding, allowing for the extraction of both explicit disciplinary labels and implicit interdisciplinary signals. This improves classification precision, adaptability across data sources, and processing efficiency.

Unlike traditional systems, ChatGPT supports multi-label classification, capturing the full disciplinary scope of publications. It also accommodates varying classification standards, reducing inconsistencies in multi-source data integration [59].

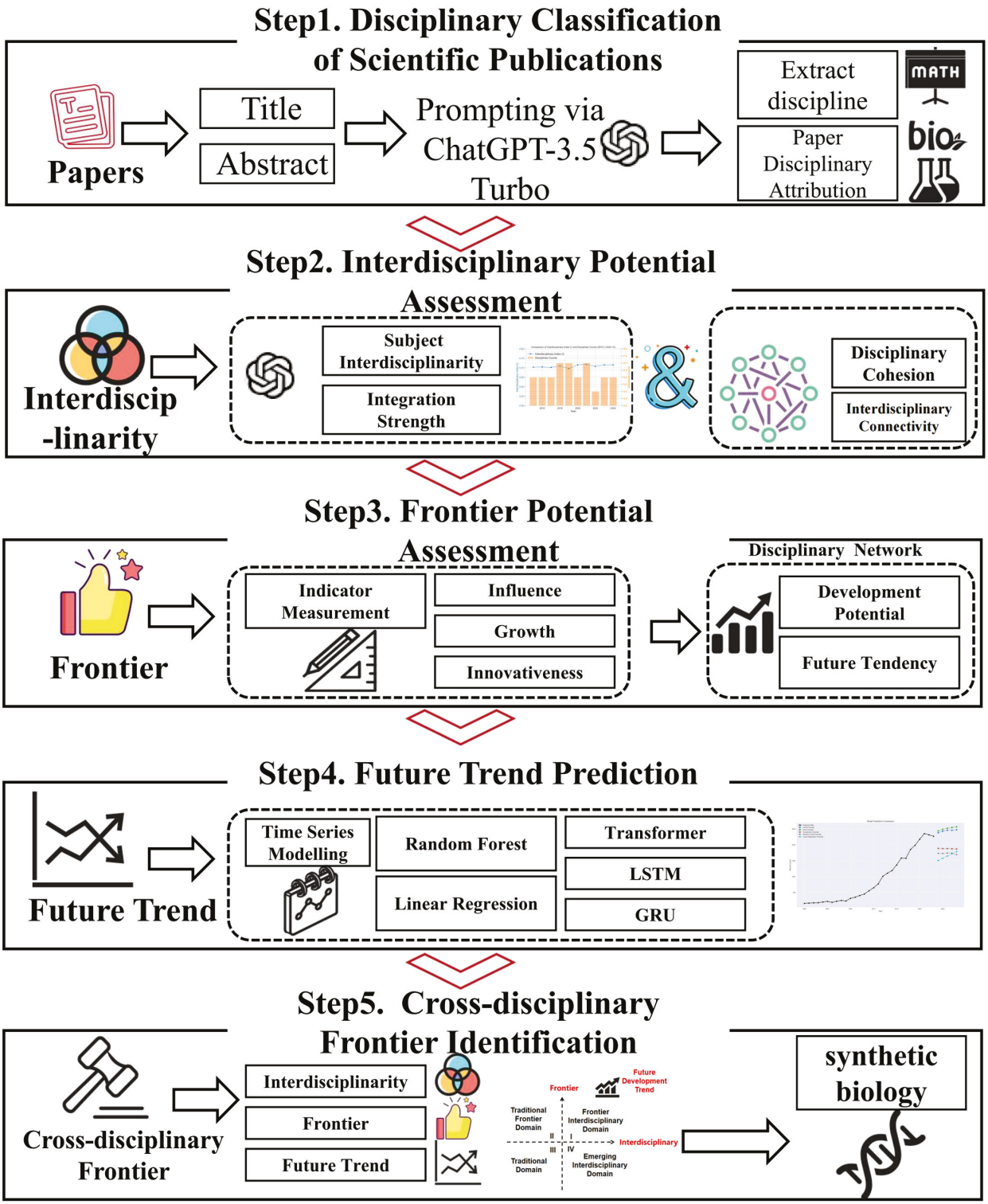


Figure 1. Research framework diagram. The framework includes five steps: (1) multi-label disciplinary classification via ChatGPT-3.5-Turbo; (2) interdisciplinarity assessment using structural and network metrics; (3) frontier potential evaluation through growth, influence, and novelty indicators; (4) trend forecasting using five time series models; and (5) identification of high-potential frontier domains through integrated analysis.

In this study, papers on synthetic biology were classified into 19 disciplines—such as Biology, Engineering, Computer Science, Medicine, and Environmental Science—based on Biglan [60] framework. This classification underpins the subsequent measurement of interdisciplinarity metrics.

To evaluate the ChatGPT-based classification system, we randomly sampled 400 articles from the full corpus. Each article was associated with a Web of Science (WOS) subject category. To establish a reliable ground truth, two domain experts independently reviewed each article's title and abstract and either confirmed the WOS label or provided a revised label. Disagreements were resolved through discussion, resulting in a consensus-based reference set.

Evaluation was conducted using standard metrics for multi-class classification:

(1) Precision: indicates the proportion of papers that were classified into a certain category by the model and indeed belong to that category. It is calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

where TP denotes true positives and FP denotes false positives.

(2) Recall: indicates the proportion of papers that actually belong to a certain category and were successfully identified by the model. It is calculated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

where FN denotes false negatives.

(3) F1 score: the harmonic mean of precision and recall, used to comprehensively assess the model's accuracy and coverage. It is calculated as follows:

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{3}$$

3.1.2. Indicators for Disciplinary Interdisciplinarity Degree and Disciplinary Integration Strength

To quantitatively assess the interdisciplinary nature of a research field, this study adopts two complementary indicators: Disciplinary Interdisciplinarity Degree (I) and Disciplinary Integration Strength (W).

The Interdisciplinarity Degree (I) reflects the breadth of disciplinary participation—that is, how evenly research output is distributed across multiple disciplines. A higher value suggests more balanced involvement from diverse fields, indicating wider interdisciplinary coverage. Based on Ma et al. [61], the formula is as follows:

$$I = 1 - \frac{\sum_{i=1}^{\gamma} m_i^2}{\left(\sum_{i=1}^{\gamma} m_i\right)^2} \tag{4}$$

This formula assesses the breadth of interdisciplinary integration within a field. A higher I value closer to 1 indicates a greater level of interdisciplinarity, meaning that the field involves a more diverse range of disciplines.

A higher interdisciplinarity score I indicates that research output is more evenly distributed across disciplines. For example, if a field includes equal contributions from biology and computer science, I will be high. In contrast, if one discipline dominates, I decreases, signaling limited interdisciplinary breadth.

The Integration Strength (W_{ij}) measures the depth of interaction between two specific disciplines by evaluating how frequently they co-occur in the same paper. A higher value indicates closer interdisciplinary collaboration. Following Leydesdorff [62], the formula is as follows:

$$W_{ij} = \frac{N(i \cap j)}{N(i \cup j)} \times 100 \tag{5}$$

where $N(i \cap j)$ represents the number of papers associated with both discipline i and discipline j, and $N(i \cup j)$ represents the total number of papers associated with either discipline i or j. A higher W value indicates a stronger interdisciplinary connection between the two disciplines.

By combining I (broad disciplinary diversity) with W (micro-level integration closeness), the framework captures both the extent and strength of interdisciplinary collaboration within a given research domain.

3.1.3. Temporal Indicators of Interdisciplinary Integration Strength

To capture the structural evolution of interdisciplinarity in synthetic biology, this study analyzes both the overall network structure and its yearly dynamics. We constructed a static disciplinary co-occurrence network based on ChatGPT-derived classifications, where nodes represent disciplines and edges reflect their co-occurrence within the same paper, weighted by frequency. This network visualizes the roles of core and peripheral disciplines and the patterns of disciplinary interaction.

To examine temporal dynamics, we divided the data into annual time windows and constructed a series of disciplinary evolution networks from 2015 to 2024. On this basis, we selected representative network topology indicators (Table 2).

Table 2. Temporal indicators of interdisciplinary integration strength.

Primary Indicator	Secondary Indicator	Indicator Type	Calculation Method
Temporal Integration Strength	Disciplinary Cohesion	Network Density	Measures the overall connectivity tightness
	Interdisciplinary Connectivity	Average Degree	Describes the average number of connections per discipline

These indicators provide insights into how disciplinary integration has changed over time, revealing key turning points and long-term collaboration trends.

3.2. Frontierness Identification of Research Fields

In the bibliometric analysis of scientific research, measuring frontierness has become an important method for identifying emerging research areas and innovation trends. Scholars have proposed various indicators for evaluating frontierness, including novelty, growth, and impact.

This study systematically characterizes the frontierness of research fields by integrating three dimensions: novelty, growth, and impact. (1) Novelty evaluates the emerging nature of research content, reflecting the speed of knowledge renewal within a field. (2) Growth measures research activity and expansion potential, indicating the developmental trends of the field. (3) Impact reflects the degree to which research outputs are cited, indicating their academic recognition and dissemination. These three dimensions collectively provide a comprehensive quantitative assessment of frontierness, capturing the degree of content innovation, the rate of field expansion, and the academic influence of research achievements.

3.2.1. Novelty Indicator

Novelty is a critical metric for determining whether the research content of a field belongs to an emerging domain. It evaluates the level of innovation within a field during a specific time period by analyzing the publication dates of papers. The calculation formula is as follows:

$$N(D) = \frac{1}{M_t(D)} \sum_{m \in M_t(D)} T_m \tag{6}$$

where $M_t(D)$ represents the number of papers in field D during time period t, and T_m denotes the publication year of paper m within the same period. A higher concentration of recent publications indicates that field D is characterized by newer research activities, suggesting a higher degree of novelty.

3.2.2. Growth Indicator

Growth reflects whether research activities in a field show an increasing trend over a specific time period. By calculating the growth rate of the number of publications across different periods, the activity level and development potential of the field can be assessed. The growth indicator is calculated as follows:

$$I(D) = \frac{1}{4}\sum_{t=1}^{4} \frac{M_{t+1}(D) - M_t(D)}{M_t(D)} \tag{7}$$

where $M_t(D)$ denotes the number of papers in field D during time period t. A higher growth value suggests that the field is rapidly developing and possesses strong potential for future expansion.

3.2.3. Impact Indicator

Impact reflects the academic influence of research outputs within a field, typically measured through citation data. The calculation formula for impact is as follows:

$$E(D) = \frac{1}{M_t(D)} \sum_{m \in M_t(D)} C_m \tag{8}$$

where C_m represents the citation count of paper m during time period t. The average number of citations per paper provides a measure of the academic influence of the field. A higher impact value indicates broader academic recognition and higher citation rates for the field's research outputs.

3.3. Field Trend Analysis Methods

3.3.1. Time Series Models

Given the temporal variability [63] and complexity [64] characteristics of frontier interdisciplinary fields, traditional static analysis methods are often insufficient for capturing the dynamic relationships among multidimensional data. In contrast, time series models can achieve both long-term trend forecasting and short-term fluctuation analysis [65], making them widely applied in the dynamic tracking and prospective research of scientific frontiers.

This study employs five representative time series models for trend analysis: LSTM [66], GRU [67], Transformer [68], Random Forest [69], and Linear Regression [70]. LSTM and GRU are selected for their strong capability in capturing temporal dependencies in sequential data; Transformer, with its attention mechanism and parallel computation, offers enhanced efficiency and scalability for long sequence modeling. Random Forest is employed for its robustness in handling nonlinear relationships, while Linear Regression serves as a baseline for performance comparison.

By integrating deep learning and traditional machine learning approaches, this study aims to comprehensively assess the dynamic trends of frontier fields under different modeling perspectives.

3.3.2. Evaluation Metrics for Time Series Models

To assess the performance of time series models in forecasting frontier trends, this study adopts five standard evaluation metrics: Mean Absolute Error (MAE), Mean Squared

Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2).

MAE and MAPE offer intuitive measures of average error; MSE and RMSE emphasize larger deviations, with RMSE reflecting error dispersion. R^2 quantifies the proportion of variance explained, indicating model fit.

Together, these indicators support a robust comparative analysis of forecasting accuracy and inform model selection in science and technology trend prediction.

3.4. Identification of Frontier Interdisciplinary Fields

To identify frontier interdisciplinary domains, this study integrates three dimensions—interdisciplinarity, frontierness, and development trend—into a unified analytical framework. Based on threshold values for each dimension, domains are classified into four types (Figure 2), enabling a structured interpretation of their innovative potential and scientific positioning.

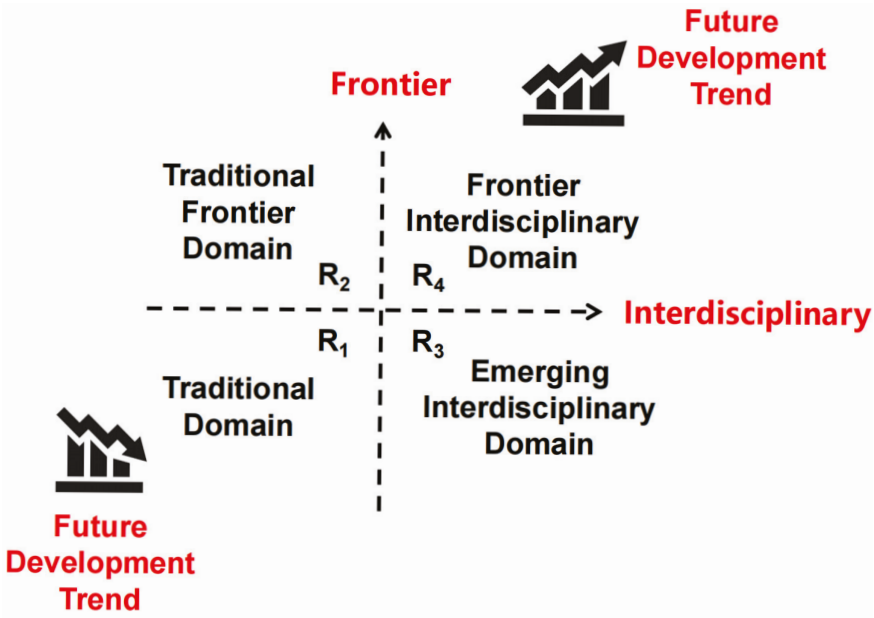


Figure 2. Framework for identifying frontier interdisciplinary fields. The diagram illustrates a classification model based on three dimensions: interdisciplinarity (horizontal axis), frontierness (vertical axis), and future development trend (indicated by arrows). The four quadrants represent distinct domain types: R_1—Traditional Domains, R_2—Traditional Frontier Domains, R_3—Emerging Interdisciplinary Domains, and R_4—Frontier Interdisciplinary Domains. This framework enables structured identification of high-potential research areas.

The conceptual function for determining frontier interdisciplinary fields is defined as follows:

$$F(x,y,z) = \begin{cases} R_1 & if \quad x < \theta_x, y < \theta_y, z < \theta_z \\ R_2 & if \quad x < \theta_x, y \geq \theta_y, z < \theta_z \\ R_3 & if \quad x \geq \theta_x, y < \theta_y, z \geq \theta_z \\ R_4 & if \quad x \geq \theta_x, y \geq \theta_y, z \geq \theta_z \end{cases} \tag{9}$$

where x: interdisciplinarity, y: frontierness, z: future potential. The thresholds θ_x, θ_y, and θ_z distinguish the following domain types:

(1) R_1—Traditional Domains: Low in all three dimensions, representing mature fields with established frameworks but limited growth and integration, such as classical thermodynamics. (2) R_2—Traditional Frontier Domains: High frontierness but low interdisciplinarity and future momentum, often focused on deepening existing core knowledge.

(3) R_3—Emerging Interdisciplinary Domains: High interdisciplinarity and future potential but low current frontierness. These areas are undergoing theoretical accumulation and cross-domain integration, e.g., agriculture + big data. (4) R_4—Frontier Interdisciplinary Domains: High across all three dimensions. These domains are deeply integrated, highly innovative, and poised for continued expansion—examples include bioinformatics, quantum computing, and AI.

The classification into R_1–R_4 research types is built on a conceptual thresholding function, using θ_x, θ_y, and θ_z to denote decision boundaries along interdisciplinarity, frontierness, and trend potential axes, respectively. In this study, we do not implement fixed numerical thresholds, but rather present a logic-based typology that can be adapted in future applications. These thresholds can be parameterized using statistical (e.g., percentile cutoffs), expert-driven, or clustering-based methods depending on specific analytical contexts.

This framework provides a scalable approach for classifying emerging research domains and supporting evidence-based science policy and information management decisions.

4. Empirical Analysis

4.1. Data Source

This study uses synthetic biology as a representative domain to validate the proposed AI-assisted framework for identifying and predicting frontier interdisciplinary fields. As a rapidly evolving area, synthetic biology integrates biology, engineering, and information science, with expanding applications in medicine, agriculture, and environmental systems.

We retrieved data from the Web of Science Core Collection, focusing on articles and reviews published between 2015 and October 2024. The SCI-EXPANDED and SSCI databases were selected for their high-quality coverage and robust citation metadata. A total of 13,683 relevant records were obtained after screening.

The 2024 data used in this study include publications up to October only, due to the timing of data retrieval. While this may lead to a slight underestimation of full-year output, it does not compromise the methodological validity of the proposed framework. On the contrary, one of the framework's key strengths is its ability to operate with partial-year or real-time data, enabling timely identification and forecasting of frontier fields without the need to wait for complete annual records.

To ensure comprehensive topic coverage, a tailored search strategy targeting key terms in synthetic biology was applied. The full query is provided in File S1.

4.2. Interdisciplinarity Identification

4.2.1. Disciplinary Classification and Interdisciplinary Scale Analysis

In this study, the ChatGPT-based literature classification system achieved strong performance across evaluation metrics. Specifically, the system reached an average precision of 0.850, recall of 0.801, and F1 score of 0.835 across the 19 disciplinary categories. These results demonstrate that the model effectively captures both explicit and implicit disciplinary signals, providing reliable multi-label classification capabilities suitable for interdisciplinary analysis. Overall, the system's performance aligns with commonly accepted benchmarks for high-quality multi-class classification tasks in literature analysis.

Building on this robust classification foundation, we analyzed the temporal evolution of disciplinary composition in synthetic biology. Figure 3 provides a longitudinal heatmap of publication counts by discipline from 2015 to 2024. The results reveal a clear trajectory of interdisciplinary convergence:

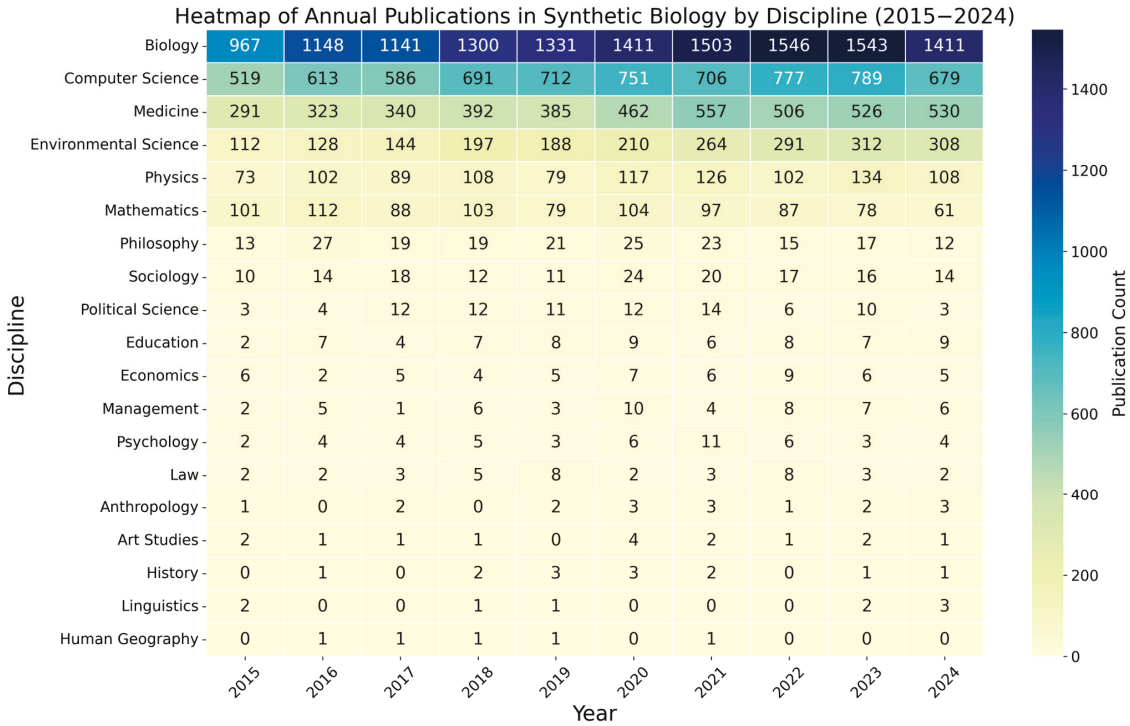


Figure 3. Annual publication heatmap showing the disciplinary composition of synthetic biology literature from 2015 to 2024. The chart illustrates the transition from a biology-dominated domain to a stabilized, multidisciplinary configuration with increasing participation from computer science, environmental science, law, and ethics.

Phase I: Foundational Phase (2015–2016)—Biology dominates the field, contributing over 60% of total publications. This reflects the early-stage emphasis on core biological processes, such as gene editing, metabolic pathways, and synthetic constructs.

Phase II: Expansion Phase (2017–2022)—A marked rise is observed in contributions from Computer Science, Environmental Science, and Engineering. The surge in Computer Science correlates with the integration of machine learning and bio-design automation. Environmental Science contributes to risk assessment and sustainable deployment, indicating the field's expanding application scope.

Phase III: Convergence Phase (2023–2024)—The disciplinary composition stabilizes, reflecting a matured interdisciplinary system. Biology remains central, but the sustained presence of Computer Science, Medicine, and Engineering points to established role-sharing: biology for foundational knowledge, CS for modeling and automation, and engineering for system integration.

Importantly, less traditional disciplines—notably Law, Ethics, and Philosophy—begin to appear more consistently post-2020. Their increasing visibility aligns with the growing recognition of biosafety, governance, and public acceptability issues, signaling a shift toward responsible innovation in synthetic biology.

These results collectively illustrate that synthetic biology is no longer a domain confined to life sciences. Instead, it is evolving into a socio-technical enterprise that increasingly draws from diverse knowledge systems, reinforcing its classification as a frontier interdisciplinary field.

Synthetic biology, as a highly interdisciplinary field, has seen notable shifts in its integration patterns over the past decade. As shown in Figure 4, interdisciplinary publications began gaining momentum in 2016, followed by rapid growth until 2022. Although publication counts declined slightly after peaking in 2022, they remained high, underscoring the sustained vitality of cross-disciplinary collaboration.

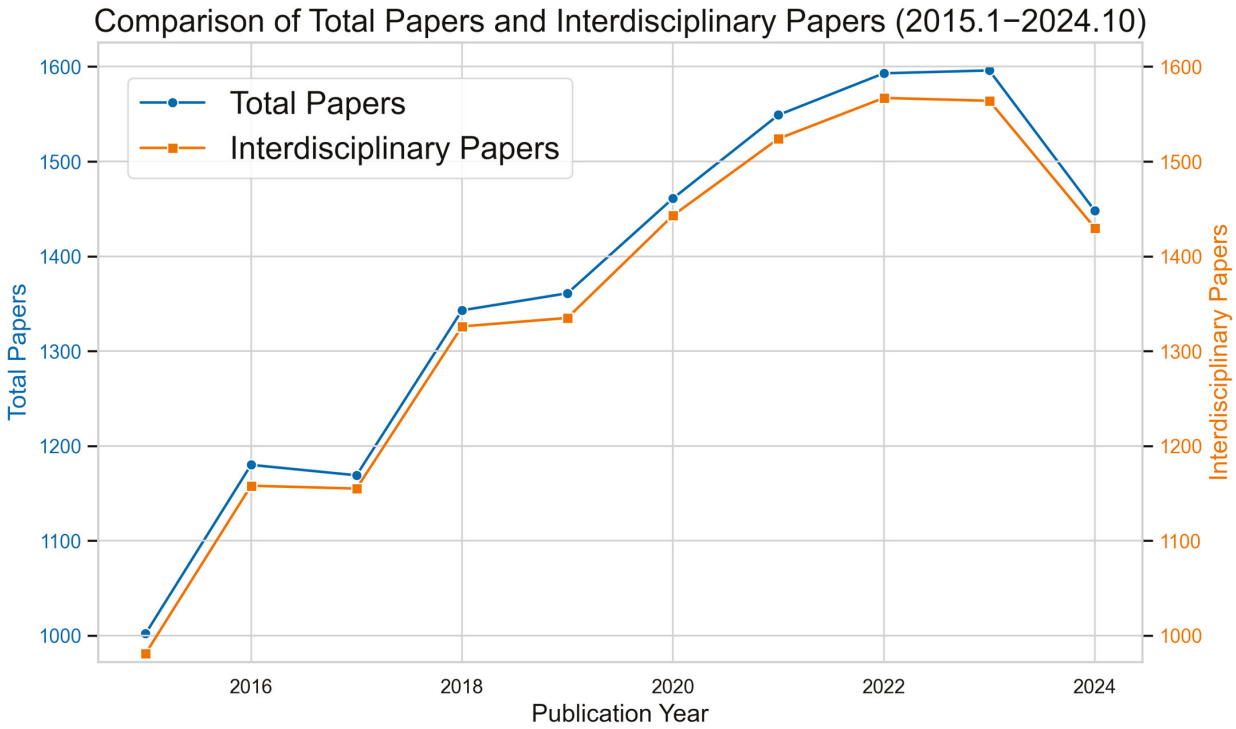


Figure 4. Proportion of interdisciplinary publications in synthetic biology (2015–2024).

The proportion of interdisciplinary papers also steadily increased from 2016 to 2022, reflecting the expanding role of external fields such as Computer Science, Engineering, and Environmental Science in synthetic biology. These disciplines have supported advances in bioinformatics, system design, and ecological assessment, respectively.

Post-2022, the proportion of interdisciplinary work remained elevated despite a drop in total publications, indicating a stabilization phase in the field's integration trajectory. Overall, synthetic biology has firmly established itself as a typical interdisciplinary domain, with continued expansion expected as new disciplines engage in its development.

4.2.2. Analysis of Interdisciplinary Integration Strength

The interdisciplinary nature of synthetic biology is reflected not only in the growth of interdisciplinary publication numbers but also in changes in interdisciplinary integration strength. Figure 5 shows the trends in the Interdisciplinary Index (I) and the number of involved disciplines (Discipline Counts) from January 2015 to October 2024.

Overall, the interdisciplinary strength of synthetic biology has exhibited relatively small fluctuations over the past decade, consistently maintaining a value around 0.7, indicating a stable level of disciplinary integration.

Specifically, the interdisciplinary strength was approximately 0.7 in 2016, and remained relatively stable with a slight upward trend from 2016 to 2018. However, a slight decline was observed between 2018 and 2020, returning to around 0.7. Since 2020, the index has stabilized again and continued through 2024.

These changes suggest that although the degree of interdisciplinarity among different disciplines may experience short-term fluctuations, overall, the integration of disciplines in synthetic biology has become structurally stable, with relatively steady collaborative relationships across fields.

Meanwhile, the trend in discipline counts also reflects the breadth of interdisciplinary integration. In 2016, approximately 16 disciplines were involved in synthetic biology research. This number increased significantly between 2016 and 2018, and stabilized around 17 disciplines during 2018–2020.

This period marked a deepening of multidisciplinary integration, as more fields became actively engaged in synthetic biology research. However, after 2022, a downward trend in discipline counts appeared, with the number falling back to about 16 disciplines by 2024. This suggests that although the breadth of interdisciplinary involvement expanded from 2016 to 2022, in recent years, the intensity of interest from certain disciplines may have declined, leading to a slight contraction in the scope of interdisciplinary engagement.

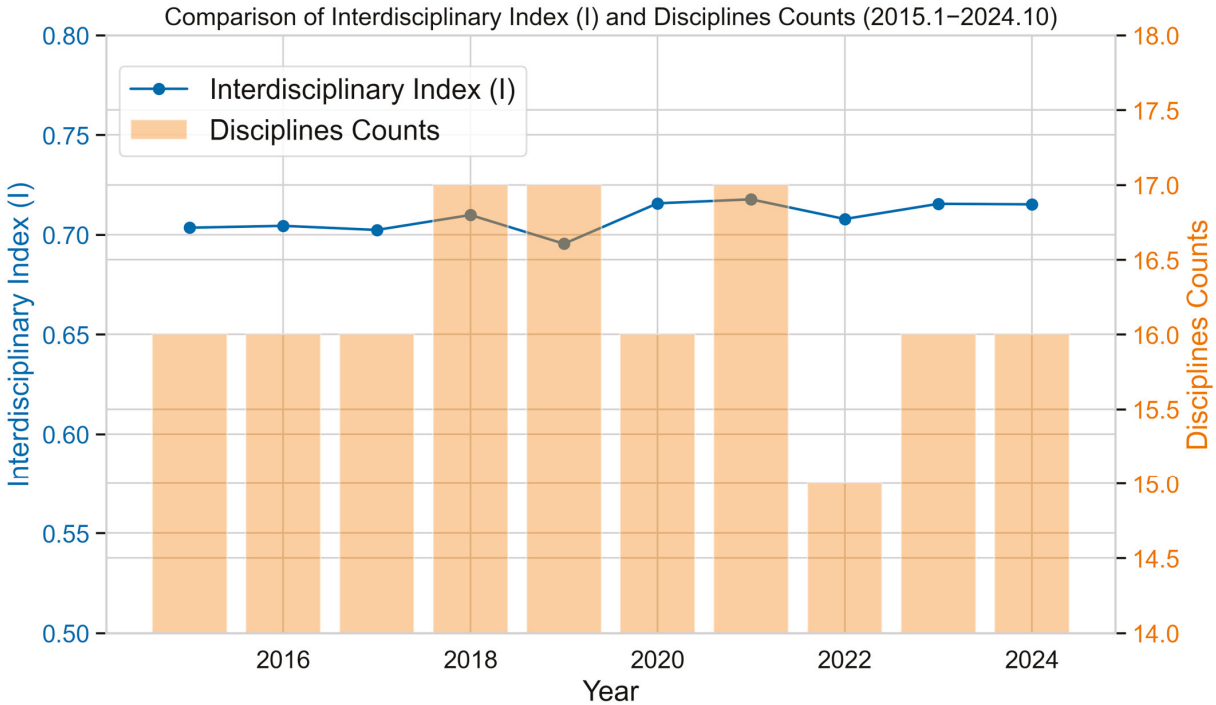


Figure 5. Trends in interdisciplinary integration strength (Interdisciplinary Index, I) and number of involved disciplines in synthetic biology from 2015 to 2024. The figure illustrates the relative stability of interdisciplinary collaboration intensity over the past decade, alongside fluctuations in the breadth of disciplinary participation.

In summary, although the number of involved disciplines has slightly decreased in recent years, the interdisciplinary integration strength has remained stable. This reflects that synthetic biology continues to maintain a high level of disciplinary fusion.

The stability of interdisciplinary strength indicates that the core disciplines have established mature and deep collaborative frameworks. Moving forward, the development of synthetic biology may rely more on the deep integration among existing disciplines rather than the large-scale influx of new disciplines.

4.2.3. Overall Network Analysis of Interdisciplinary Connections

Figure 6 visualizes the disciplinary co-occurrence network in synthetic biology, highlighting the structural characteristics of interdisciplinary collaboration in the field.

At the core of the network is biology, which holds the largest node and highest connectivity. This reflects its foundational role and its extensive interaction with other disciplines.

Computer science is closely linked to biology, driven by the rise of computational biology, bioinformatics, and AI-assisted research. Environmental science and sociology also maintain strong ties with biology, contributing to ecological risk assessments and societal impact evaluations.

Other disciplines such as medicine, engineering, and neuroscience connect through direct or intermediate paths. Linguistics, linked via computer science, reflects the growing role of natural language processing in biological knowledge mining.

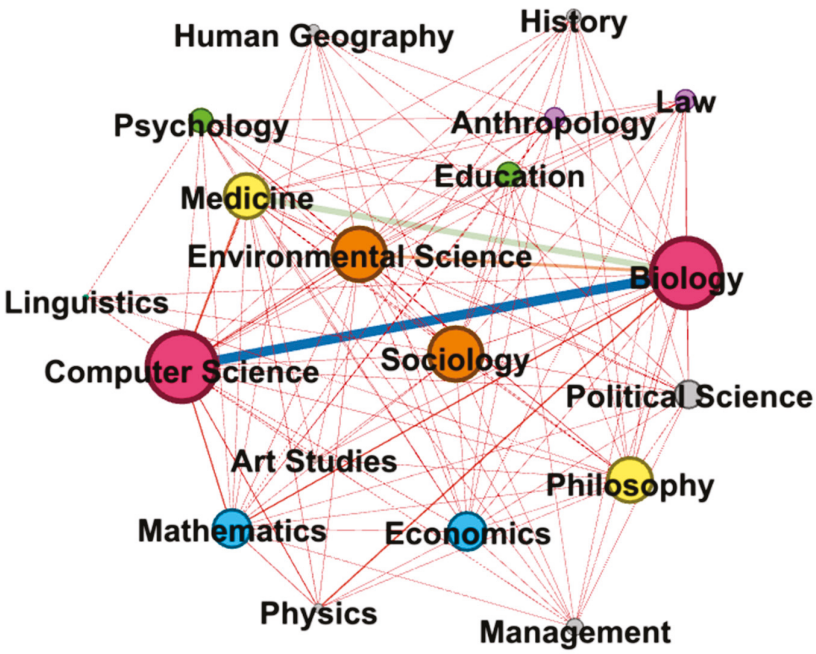


Figure 6. Interdisciplinary co-occurrence network of synthetic biology publications (2015–2024). Node size represents publication volume; edge thickness reflects co-occurrence frequency. The network reveals a biology-centered structure with strong computational integration and emerging governance-oriented peripheries.

Peripheral fields like law, philosophy, and political science appear less frequently but remain important in shaping ethical frameworks, regulatory policies, and public engagement. In the interdisciplinary structure of synthetic biology, the presence of disciplines such as ethics, law, and public policy is notable. These fields contribute to shaping the research landscape in important ways. Ethical debates on genome editing, regulatory frameworks for biosafety, and legal mechanisms for intellectual property all affect funding directions, institutional behavior, and even the public legitimacy of emerging technologies. Their appearance in the citation network and semantic classification suggests that frontier science is increasingly embedded within broader societal dialogues, where governance structures co-evolve with technical innovation. This co-evolutionary dynamic deserves further attention in future studies of frontier field development.

This network structure provides three key insights: (1) Synthetic biology is not merely interdisciplinary—it is polycentric, involving diverse epistemic communities that contribute at different layers (technical, methodological, institutional). (2) Interdisciplinary integration is asymmetric: biology collaborates intensively with computational and environmental fields, while legal, philosophical, and political domains maintain functional but looser ties. (3) The presence of peripheral social sciences signals an ongoing shift toward responsible and anticipatory governance, which will likely intensify as applications of synthetic biology broaden.

Overall, the network demonstrates a biology-centered, multi-level integration structure, with core disciplines enabling intensive collaboration and peripheral ones providing institutional and ethical context. As synthetic biology advances, deeper interdisciplinary expansion is expected.

4.2.4. Evolutionary Analysis of the Interdisciplinary Network

The evolution of interdisciplinary research in synthetic biology is reflected in the structural dynamics of its disciplinary co-occurrence network (Figure 7). Key topological metrics from 2015 to 2024 (Table S1) reveal distinct shifts in connectivity and collaboration intensity over the past decade.

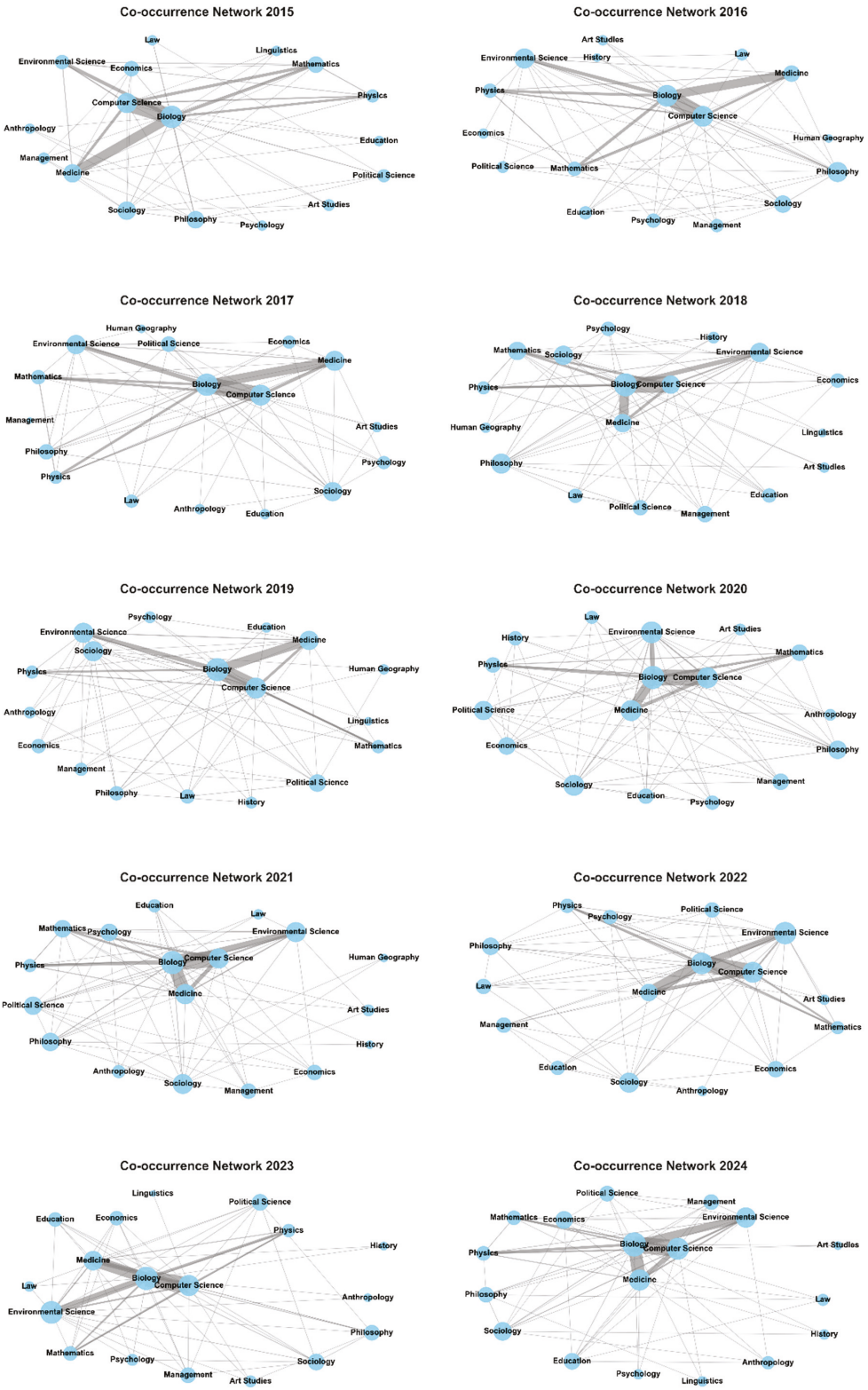


Figure 7. Disciplinary evolution network of synthetic biology (2015–2024).

From 2015 to 2019, network density remained relatively stable (0.40–0.44), indicating consistent interdisciplinary linkages. In 2020, however, both density and average degree rose sharply—reaching 0.57 and 9.06 respectively—suggesting intensified disciplinary interaction, likely catalyzed by the global COVID-19 crisis. While these values slightly declined in subsequent years, they remained elevated, pointing to a lasting structural shift.

The clustering coefficient remained high throughout the period (0.76–0.81), implying stable subgroup formation and strong internal cohesion within the interdisciplinary network. Meanwhile, the network maintained a short diameter (2–3) and average path length (1.43–1.62), reflecting efficient disciplinary connectivity.

The number of disciplines and their interactions also increased, expanding from 17 nodes and 55 edges in 2015 to 18 nodes and 68 edges by 2024. This growth highlights not only the diversification of participating fields but also the intensification of interdisciplinary collaboration.

Overall, 2020 marked a structural turning point. Despite modest fluctuations thereafter, synthetic biology has sustained high levels of integration and connectivity, underscoring its status as a mature and enduringly interdisciplinary field.

4.3. Frontierness Analysis of the Field

Synthetic biology demonstrates strong frontier characteristics, reflected in its novelty, growth, and impact indicators.

4.3.1. Novelty

Based on Equation (6), the novelty score of synthetic biology is 2019.860, indicating that research hotspots in the field are highly concentrated in recent years, reflecting strong temporal novelty.

This is primarily attributed to breakthroughs in key technologies, such as CRISPR gene editing, automated laboratory platforms, and AI-assisted synthetic pathway optimization, which have continually pushed the research frontiers forward. The high novelty score suggests that synthetic biology remains in a phase of rapid evolution, with the potential for further technological and application breakthroughs in the future (Table 3).

Table 3. Novelty scores of some research areas.

Research Area	Novelty Score
Synthetic Biology	2019.860
Information Technology and Services	2006.220
Algorithm Optimization	2011.191
Medical Data Modeling	2021.260
Machine Learning and Deep Learning	2021.219

Compared to other research topics, synthetic biology's novelty score is significantly higher than that of Information Technology and Services (2006.220) and Algorithm Optimization (2011.191), and is close to the scores for emerging cross-disciplinary technologies such as Medical Data Modeling (2021.260) and Machine Learning and Deep Learning (2021.219) [71], highlighting its position at the forefront of research.

4.3.2. Growth

Based on Equation (7), the growth rate of synthetic biology is calculated as 4.46%, indicating steady research expansion. Although its growth rate is not as explosive as that of emerging technologies such as artificial intelligence, it maintains a healthy upward trend. Such stable growth typically suggests that the field has entered a sustainable development phase, supported by a stable research community and a gradually maturing technological system. This trend is closely linked to the application-driven nature of synthetic biology,

with advances in biomanufacturing, precision medicine, and agricultural biotechnology driving the steady translation of research results into practical applications.

4.3.3. Impact

Based on Equation (8), the impact score of synthetic biology is 25.03, positioning it at a moderate-to-high level among various research fields. Although slightly lower than Information Technology and Services (55.73) and Scientometrics (43.08), it is notably higher than Medical Data Modeling (10.73) and Machine Learning and Deep Learning (11.12) [71]. This indicates that synthetic biology has achieved considerable academic influence, policy support, and patent activity.

The strong impact of the field is largely due to its critical position within a multidisciplinary context, particularly through its deep integration with fields such as biomedicine, agricultural engineering, and environmental remediation, thus enabling its research outcomes to generate broad and lasting value in real-world applications.

4.3.4. Field Potential

Synthetic biology is fundamentally a technology- and application-driven discipline, propelled by advances in gene editing, bioengineering methodologies, and computational simulation technologies. In recent years, the maturation of tools like CRISPR has contributed significantly to the simultaneous growth in publication output and academic influence within the field. The field's interdisciplinary nature is another crucial factor supporting its continuous expansion. By integrating Biology, Engineering, Computer Science, and other disciplines, synthetic biology continually explores new research directions and enhances its academic impact.

Based on the above analyses, synthetic biology exhibits substantial potential for future growth, particularly in cross-disciplinary applications such as AI-assisted gene design, computational modeling for pathway optimization, and intelligent biomanufacturing processes.

Although the current growth rate (4.46%) is relatively stable, the field's high impact (25.03) and high frontierness (2019.860) indicate the high quality of ongoing research, suggesting that more breakthrough innovations are likely to emerge. Furthermore, as industry interest in synthetic biology continues to rise, the trend toward integrating academic research with industrial applications is expected to strengthen, providing additional momentum for field development. Consequently, synthetic biology is likely to maintain strong frontier characteristics and exert profound scientific and societal impacts across multiple domains in the future.

4.4. Trend Analysis of the Field

4.4.1. Model Construction and Error Evaluation

To forecast synthetic biology's research trajectory, five models were implemented: three deep learning models (LSTM, GRU, Transformer) and two baseline models (Random Forest and Linear Regression). All models were trained using TensorFlow 2.6.0 and Keras 2.9.0, and evaluated using standard error metrics (MAE, MSE, RMSE, MAPE, R^2).

As shown in Figure 8, deep learning models outperformed traditional ones. The Transformer exhibited the lowest and most stable error values, suggesting strong generalization performance. In contrast, LSTM and GRU showed fluctuating MAPE curves, indicating potential instability. Random Forest and Linear Regression failed to capture the underlying nonlinear temporal dynamics.

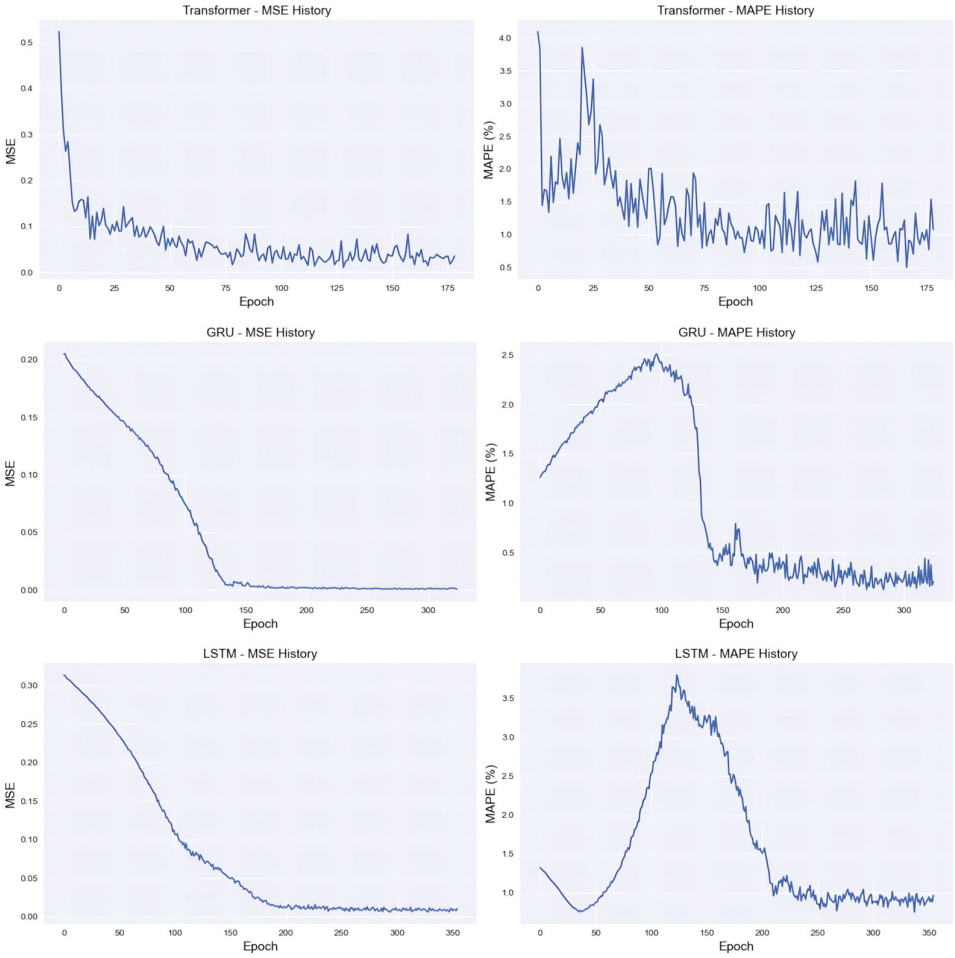


Figure 8. Convergence curves comparison of deep learning temporal models (Transformer, GRU, LSTM).

4.4.2. Model Comparison and Best Model Selection

Table 4 summarizes the prediction performance of all models on the test set using five metrics: Mean Absolute Error (MAE), MSE, Root Mean Squared Error (RMSE), MAPE, and R^2. Among all models, the Transformer achieved the best overall performance, with the lowest error values across all indicators and a high R^2 of 0.96, indicating excellent predictive accuracy and data fitting.

Table 4. Performance comparison of forecasting models.

Model	MAE	MSE	RMSE	MAPE/%	R^2
LSTM	0.21	0.09	0.30	1.52	0.23
GRU	0.35	0.19	0.43	2.42	0.00
Transformer	0.06	0.00	0.07	0.22	0.96
Random Forest	0.58	0.45	0.67	8.85	0.00
Linear Regression	0.62	0.47	0.69	8.57	0.00

The LSTM ranked second with relatively low error values but showed a limited capacity for generalization (R^2 = 0.23). GRU exhibited higher error rates, particularly underperforming in MAPE and R^2. Traditional models such as Random Forest and Linear Regression failed to capture the nonlinear patterns inherent in synthetic biology trends, performing poorly across all metrics.

To further validate model effectiveness, Figure 9 compares historical publication data with forecasts from each model. The Transformer aligned most closely with the observed slowdown in growth after 2022, projecting steady expansion rather than an explosive increase. In contrast, LSTM predicted exponential growth, likely overfitting early acceleration trends. GRU showed more conservative forecasts, while Random Forest and Linear Regression projected near-flat growth, clearly underestimating the field's dynamics. These results underscore the Transformer's superior ability to capture complex temporal patterns in emerging research domains.

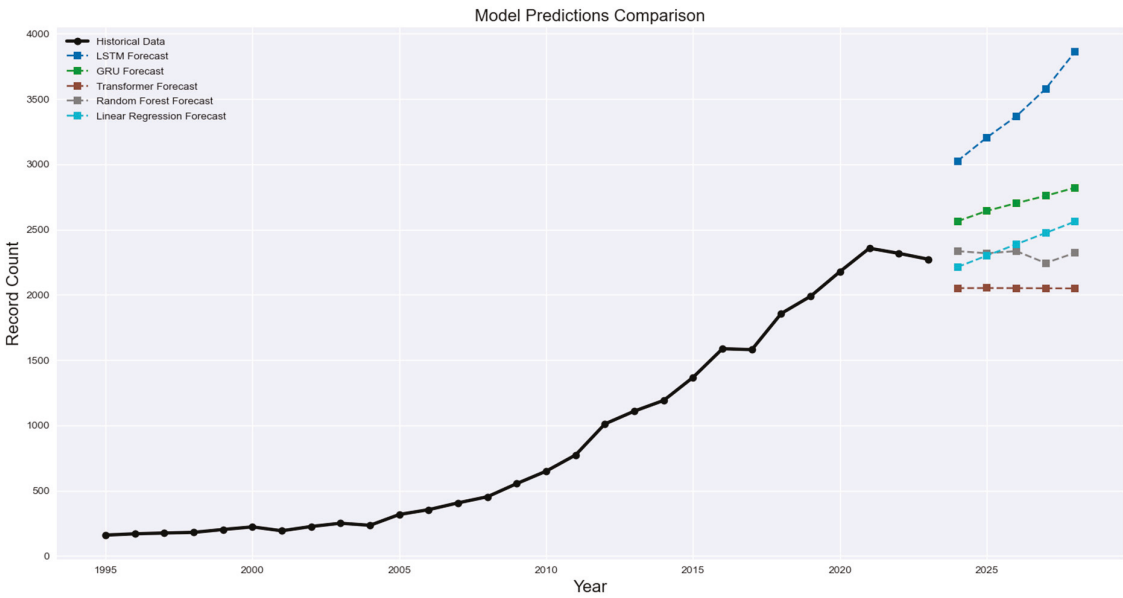


Figure 9. Model predictions comparison across temporal forecasting techniques. The Transformer model best fits the recent publication plateau and predicts steady future growth. In contrast, LSTM forecasts an exponential rise, GRU remains conservative, while traditional models underestimate the trend—visually confirming the Transformer's superiority.

4.5. Identification of Frontier Interdisciplinary Domain

Based on the interdisciplinarity–frontierness–trend framework proposed in this study (Figure 2), synthetic biology can be identified as a typical frontier interdisciplinary domain.

4.5.1. Interdisciplinarity and Frontierness

Synthetic biology integrates knowledge from biology, computer science, engineering, environmental science, and other disciplines, reflecting strong interdisciplinary collaboration. Research in areas such as AI-assisted gene design, computational biology, and biomanufacturing has reinforced its role as a convergence hub for diverse technologies.

From a frontierness perspective, synthetic biology demonstrates both novelty and impact: it features recent high-intensity research activity and increasing academic and industrial attention around topics like CRISPR, synthetic circuits, and artificial cells. These indicators confirm its status as a highly innovative and scientifically forward-looking field.

4.5.2. Trend Outlook and Strategic Role

Since 2010, synthetic biology has entered a phase of accelerated growth, peaking around 2020. Forecasting results—especially from Transformer models—suggest sustained upward momentum, with new opportunities emerging at the intersection of synthetic biology and artificial intelligence. Applications in agriculture, medicine, and environmental engineering are also expanding, reinforcing the domain's practical relevance.

4.5.3. Final Positioning

Synthesizing findings across all three dimensions, synthetic biology clearly qualifies as a frontier interdisciplinary domain. It combines a high degree of disciplinary integration with robust innovation capacity and stable development potential. As such, it is well positioned to drive cross-disciplinary technological transformation and contribute strategically to the future landscape of science, policy, and industry.

5. Discussion

This study constructs an integrated framework combining large language models, interdisciplinary metrics, and time series forecasting to identify and predict frontier interdisciplinary domains. By applying this framework to synthetic biology, we demonstrate its capability in extracting disciplinary patterns, measuring frontierness, and forecasting future trajectories. The discussion unfolds from both methodological and domain-specific perspectives.

5.1. Effectiveness of the Methodological Framework and Forecasting Algorithms

In the context of information management and science foresight, accurately identifying disciplinary frontiers and anticipating their development is a growing priority [72]. Our proposed framework addresses this need by combining semantic classification, network-based interdisciplinary indicators, and predictive modeling.

Empirical validation shows that the framework can differentiate disciplinary evolution types and offers scalable tools for research monitoring and strategic planning. Although synthetic biology is used as a test case, the methodology is adaptable to other domains, supporting broader policy and research agenda-setting.

Among the forecasting models compared, the Transformer significantly outperformed LSTM, GRU, Random Forest, and Linear Regression in both short- and long-term trend prediction, confirming its value in modeling complex nonlinear dynamics [73].

5.2. Interdisciplinary Integration and Frontier Characteristics of Synthetic Biology

Synthetic biology exemplifies a frontier interdisciplinary field, bringing together biology, computer science, engineering, environmental science, and medicine. Recent advances have driven the field toward greater automation and intelligence [74], with rapid progress in AI-assisted gene design [75] and automated biomanufacturing [76]. Its integration with environmental and medical sciences has enabled impactful applications in ecological restoration, precision medicine, and bioenergy. From a frontier perspective, the field shows high novelty and wide academic recognition. Research hotspots such as CRISPR, synthetic circuits, and artificial cells highlight its leadership in technological innovation.

Overall, synthetic biology demonstrates strong interdisciplinarity and transformative potential, reinforcing its position as a priority domain in science policy and innovation management.

5.3. Development Trends and Future Growth Forecast

Historical data and model forecasts show that synthetic biology entered a rapid growth phase after 2010, peaking around 2020. In the past two years, however, growth has slightly slowed—possibly due to resource constraints, policy regulations, or technical bottlenecks. According to the Transformer-based forecast, the field is expected to continue steady expansion, though the growth rate may stabilize, suggesting a transition into a phase of technological refinement and application-oriented expansion. Additionally, based on the frontier-interdisciplinarity-trend classification framework, future research hotspots in synthetic biology may continue to extend toward multi-domain integration.

The convergence of synthetic biology with areas such as artificial intelligence, biomedical engineering, and green technology is likely to reshape its innovation ecosystem, reinforcing its strategic role in both academic and industrial contexts.

5.4. Challenges, Limitations, and Future Directions

Despite its growing influence, synthetic biology faces challenges in technology development, ethical governance, and industrial translation. Technologically, core capabilities such as gene editing, synthetic pathway optimization, and artificial cell construction remain limited in stability, scalability, and reproducibility [77,78]. While AI has begun to assist gene design and modeling [79], enhancing biological interpretability and experimental feasibility remains a key bottleneck [80].

Ethical and biosafety concerns are intensifying as synthetic biology advances. Issues surrounding artificial life and genetic modification require stronger governance in biosafety, ethics, and ecological risk [81–83].

On the policy side, industrial adoption faces regulatory fragmentation, cost barriers, and slow market translation [84]. Harmonizing global standards and ethical frameworks is urgently needed [84–86].

GPT-3.5-Turbo was selected for its accessibility and stability. Despite its strong classification performance, challenges such as hallucinations, prompt sensitivity, and interpretability persist. Future work could explore newer models (e.g., GPT-4, DeepSeek, Gemini) for improved robustness, reasoning, and domain adaptability.

Despite its strong performance, the ChatGPT-based classification system faces concerns regarding hallucinations [87], prompt sensitivity [88], and interpretability. In some cases, the model generated field labels not present in standard taxonomies (e.g., WOS), reflecting its tendency to infer or "hallucinate [89]" plausible categories. Moreover, slight variations in prompt wording can lead to inconsistent outputs, particularly for interdisciplinary abstracts. Although temperature was fixed at 0 to reduce output randomness [90], such behaviors highlight the need for robustness testing, ensemble prompting, and possibly fine-tuning for future implementations. We consider these factors important for responsible use of LLMs in scientific evaluation workflows.

While the Transformer model demonstrated superior forecasting accuracy based on standard metrics (MAE, MSE, RMSE, R^2), its practical utility in supporting strategic foresight also depends on robustness to data perturbation. In real-world settings, historical research trajectories may be noisy, incomplete, or distorted by lagging indicators (e.g., delays in publication or indexing [91]). A small shift in forecasted growth curves—especially near inflection points—could influence funding strategies, institutional planning, or technology prioritization.

Although this study did not include a formal sensitivity analysis, future work will explore the stability of forecasting outputs under simulated data noise, varying time horizons, and missing data. Such robustness checks are essential to ensure responsible deployment of trend analysis tools in policy or investment contexts.

This study, while offering a robust framework, is limited by its reliance on bibliometric data and assumptions inherent to the modeling approaches. Future work should incorporate multi-source datasets—including patents, policy texts, and industrial records—to improve the precision of trend forecasts and assess interdisciplinary spillovers. Expanding the framework to other emerging domains would further validate its generalizability and enhance its utility for information management and policy intelligence.

The results demonstrate that our integrated framework not only enables the identification of interdisciplinary research areas but also facilitates their dynamic monitoring and future projection. Compared with traditional methods such as LDA-based topic modeling

and citation burst detection, our framework incorporates semantic classification, multidimensional indicators, and deep learning models, which together offer a more robust and forward-looking approach. Notably, the Transformer model showed superior accuracy and stability in forecasting publication trends, likely due to its ability to capture long-range dependencies and nonlinear growth patterns. The rising indicator curves of interdisciplinarity, novelty, and growth in synthetic biology further validate its status as a frontier interdisciplinary domain. This triangulation across multiple dimensions improves both the reliability and interpretability of the results, supporting data-driven decision-making in research, policy, and innovation management.

6. Conclusions

This study proposed an integrated framework for identifying and predicting frontier interdisciplinary fields by combining large language model-based semantic classification, multidimensional indicator evaluation, and time series forecasting. Using synthetic biology as a case study, we demonstrated that the framework effectively captures both the structural and temporal characteristics of emerging domains.

Key findings include: the following (1) ChatGPT-based disciplinary classification enables flexible, high-granularity field labeling; (2) the multidimensional evaluation system—comprising interdisciplinarity, novelty, growth, impact, and trend—provides a comprehensive lens for characterizing research frontiers; and (3) the Transformer model outperforms other forecasting methods in predicting future publication trends, confirming its suitability for modeling nonlinear scientific growth.

The proposed framework offers methodological advancements in science mapping, with practical implications for research management, strategic funding, and innovation policy. Future work could extend the model by incorporating additional data modalities (e.g., patents, policy documents), testing on other emerging fields, and integrating multimodal large language models for deeper semantic analysis.

Future research will apply the framework to additional interdisciplinary domains (e.g., quantum computing, bioinformatics) to further assess its generalizability and scalability.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/systems13080677/s1, Supplementary File S1: Search Query Used in the Web of Science Core Collection; Table S1: Evolution of Key Topological Indicators in the Synthetic Biology Interdisciplinary Network (2015–2024).

Author Contributions: Y.W.: Conceptualization, Methodology, Writing—original draft. R.Y.: Investigation, Data curation. J.W.: Writing—Review and Editing. Q.L.: Writing—Review and Editing. X.Z.: Methodology, Funding acquisition, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Office for Philosophy and Social Sciences, Beijing, China [Grant number 23&ZD225 2023].

Data Availability Statement: The data that support the findings of this study are available upon request to the corresponding author.

Acknowledgments: This study used ChatGPT for linguistic embellishment to improve the readability of the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Klein, J.T. *Beyond Interdisciplinarity: Boundary Work, Communication, and Collaboration*; Oxford University Press: Oxford, UK, 2021.

2. Zahra, S.A.; Newey, L.R. Maximizing the impact of organization science: Theory-building at the intersection of disciplines and/or fields. *J. Manag. Stud.* **2009**, *46*, 1059–1075. [CrossRef]
3. Keohane, R.O.; Victor, D.G. The regime complex for climate change. *Perspect. Politics* **2011**, *9*, 7–23. [CrossRef]
4. Newell, P. Globalization and the governance of biotechnology. *Glob. Environ. Politics* **2003**, *3*, 56–71. [CrossRef]
5. Dwivedi, Y.K.; Hughes, L.; Kar, A.K.; Baabdullah, A.M.; Grover, P.; Abbas, R.; Andreini, D.; Abumoghli, I.; Barlette, Y.; Bunker, D. Climate change and COP26: Are digital technologies and information management part of the problem or the solution? An editorial reflection and call to action. *Int. J. Inf. Manag.* **2022**, *63*, 102456. [CrossRef]
6. Le, T.; Pham, H.; Mai, S.; Vu, N. Frontier academic research, industrial R&D and technological progress: The case of OECD countries. *Technovation* **2022**, *114*, 102436. [CrossRef]
7. Ye, G.; Wang, C.; Wu, C.; Peng, Z.; Wei, J.; Song, X.; Tan, Q.; Wu, L. Research frontier detection and analysis based on research grants information: A case study on health informatics in the US. *J. Informetr.* **2023**, *17*, 101421. [CrossRef]
8. Huang, M.-H.; Chang, C.-P. Detecting research fronts in OLED field using bibliographic coupling with sliding window. *Scientometrics* **2014**, *98*, 1721–1744. [CrossRef]
9. Morss, R.E.; Lazrus, H.; Demuth, J.L. The "inter" within interdisciplinary research: Strategies for building integration across fields. *Risk Anal.* **2021**, *41*, 1152–1161. [CrossRef]
10. Newman, J. Promoting interdisciplinary research collaboration: A systematic review, a critical literature review, and a pathway forward. *Soc. Epistemol.* **2024**, *38*, 135–151. [CrossRef]
11. Yu, M.; Li, J.; Cui, Y. Review on Science and Technology Trend Awareness Research. In Proceedings of the 2024 3rd International Conference on Big Data, Information and Computer Network (BDICN), Sanya, China, 12–14 January 2024; pp. 1–6.
12. Ozsoy, C.M.; Mengüç, M.P. A transdisciplinary approach and design thinking methodology: For applications to complex problems and energy transition. *World* **2024**, *5*, 119–135. [CrossRef]
13. Li, T.; Cui, L.; Lv, W.; Song, X.; Cui, X.; Tang, L. Exploring the frontiers of sustainable livelihoods research within grassland ecosystem: A scientometric analysis. *Heliyon* **2022**, *8*, e10704. [CrossRef] [PubMed]
14. Amado, A.; Cortez, P.; Rita, P.; Moro, S. Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *Eur. Res. Manag. Bus. Econ.* **2018**, *24*, 1–7. [CrossRef]
15. Funk, P.; Davis, A.; Vaishnav, P.; Dewitt, B.; Fuchs, E. Individual inconsistency and aggregate rationality: Overcoming inconsistencies in expert judgment at the technical frontier. *Technol. Forecast. Soc. Change* **2020**, *155*, 119984. [CrossRef]
16. Azoulay, P.; Lynn, F.B. Self-citation, cumulative advantage, and gender inequality in science. *Sociol. Sci.* **2020**, *7*, 152–186. [CrossRef]
17. Hall, B.H.; Jaffe, A.B.; Trajtenberg, M. *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools*; National Bureau of Economic Research: Cambridge, MA, USA, 2001.
18. Gerlach, M.; Peixoto, T.P.; Altmann, E.G. A network approach to topic models. *Sci. Adv.* **2018**, *4*, eaaq1360. [CrossRef]
19. Więckowski, J.; Kizielewicz, B.; Wątróbski, J.; Sałabun, W. A New Approach for Handling Uncertainty of Expert Judgments in Complex Decision Problems. *IEEE Access* **2024**, *12*, 142026–142046. [CrossRef]
20. Lesnikowski, A.; Belfer, E.; Rodman, E.; Smith, J.; Biesbroek, R.; Wilkerson, J.D.; Ford, J.D.; Berrang-Ford, L. Frontiers in data analytics for adaptation research: Topic modeling. *Wiley Interdiscip. Rev. Clim. Change* **2019**, *10*, e576. [CrossRef]
21. Wang, X.; Zhang, S.; Liu, Y. ITGInsight–discovering and visualizing research fronts in the scientific literature. *Scientometrics* **2022**, *127*, 6509–6531. [CrossRef]
22. Wang, X.; Zhang, S.; Liu, Y.; Du, J.; Huang, H. How pharmaceutical innovation evolves: The path from science to technological development to marketable drugs. *Technol. Forecast. Soc. Change* **2021**, *167*, 120698. [CrossRef]
23. Yu, D.; Yan, Z. Combining machine learning and main path analysis to identify research front: From the perspective of science-technology linkage. *Scientometrics* **2022**, *127*, 4251–4274. [CrossRef]
24. Price de Solla, D.J. Networks of scientific papers. *Science* **1965**, *149*, 510–515. [CrossRef]
25. Small, H.; Griffith, B.C. The structure of scientific literatures I: Identifying and graphing specialties. *Sci. Stud.* **1974**, *4*, 17–40. [CrossRef]
26. Shibata, N.; Kajikawa, Y.; Takeda, Y.; Matsushima, K. Comparative study on methods of detecting research fronts using different types of citation. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 571–580. [CrossRef]
27. Williamson, K. Chapter 12—The Delphi method. In *Research Methods for Students, Academics and Professionals*, 2nd ed.; Williamson, K., Bow, A., Burstein, F., Darke, P., Harvey, R., Johanson, G., McKemmish, S., Oosthuizen, M., Saule, S., Schauder, D., et al., Eds.; Chandos Publishing: Oxfordshire, UK, 2002; pp. 209–220. [CrossRef]
28. Osareh, F. Bibliometrics, citation analysis and co-citation analysis: A review of literature I. *Libri* **1996**, *46*, 217–225. [CrossRef]
29. Weinberg, B.H. Bibliographic coupling: A review. *Inf. Storage Retr.* **1974**, *10*, 189–196. [CrossRef]
30. Zhang, T.; Chen, J.; Lu, Y.; Yang, X.; Ouyang, Z. Identification of technology frontiers of artificial intelligence-assisted pathology based on patent citation network. *PLoS ONE* **2022**, *17*, e0273355. [CrossRef] [PubMed]

31. Xu, D.; Liu, B.; Wang, J.; Zhang, Z. Bibliometric analysis of artificial intelligence for biotechnology and applied microbiology: Exploring research hotspots and frontiers. *Front. Bioeng. Biotechnol.* **2022**, *10*, 998298. [CrossRef]
32. Chen, X.; Mao, J.; Li, G. A co-citation approach to the analysis on the interaction between scientific and technological knowledge. *J. Informetr.* **2024**, *18*, 101548. [CrossRef]
33. Ma, C.; Xu, Q.; Li, B. Comparative study on intelligent education research among countries based on bibliographic coupling analysis. *Libr. Hi Tech* **2022**, *40*, 786–804. [CrossRef]
34. Yanhui, S.; Lijuan, W.; Junping, Q. A comparative study of first and all-author bibliographic coupling analysis based on Scientometrics. *Scientometrics* **2021**, *126*, 1125–1147. [CrossRef]
35. Liu, J.S.; Lu, L.Y.; Lu, W.-M. Research fronts in data envelopment analysis. *Omega* **2016**, *58*, 33–45. [CrossRef]
36. Wei, Y.-M.; Wang, J.-W.; Chen, T.; Yu, B.-Y.; Liao, H. Frontiers of low-carbon technologies: Results from bibliographic coupling with sliding window. *J. Clean. Prod.* **2018**, *190*, 422–431. [CrossRef]
37. Kleinberg, J. Bursty and hierarchical structure in streams, data mining and knowledge discovery. In Proceedings of the Elected Papers from the 8th ACM SIGKDD International Conference on Knowledge I Discovery and Data Mining? Edmonton, AB, Canada, 23–26 July 2002; pp. 372–397.
38. Li, M.; Chu, Y. Explore the research front of a specific research theme based on a novel technique of enhanced co-word analysis. *J. Inf. Sci.* **2017**, *43*, 725–741. [CrossRef]
39. Xie, T.; Qin, P.; Yan, J. Research on Artificial Intelligence Frontier Recognition Based on LDA. *Open Access Libr. J.* **2018**, *5*, 1–13. [CrossRef]
40. Wu, Q.; Kuang, Y.; Hong, Q.; She, Y. Frontier knowledge discovery and visualization in cancer field based on KOS and LDA. *Scientometrics* **2019**, *118*, 979–1010. [CrossRef]
41. Yu, D.; Xiang, B. Discovering topics and trends in the field of Artificial Intelligence: Using LDA topic modeling. *Expert Syst. Appl.* **2023**, *225*, 120114. [CrossRef]
42. Liu, Y.; Wan, F. Unveiling temporal and spatial research trends in precision agriculture: A BERTopic text mining approach. *Heliyon* **2024**, *10*, e36808. [CrossRef]
43. He, J.; Chen, C. Predictive effects of novelty measured by temporal embeddings on the growth of scientific literature. *Front. Res. Metr. Anal.* **2018**, *3*, 9. [CrossRef]
44. Park, I.; Yoon, B. Identifying promising research frontiers of pattern recognition through bibliometric analysis. *Sustainability* **2018**, *10*, 4055. [CrossRef]
45. Wang, Z.; Chen, J.; Chen, J.; Chen, H. Identifying interdisciplinary topics and their evolution based on BERTopic. *Scientometrics* **2024**, *129*, 7359–7384. [CrossRef]
46. Xu, J.; Bu, Y.; Ding, Y.; Yang, S.; Zhang, H.; Yu, C.; Sun, L. Understanding the formation of interdisciplinary research from the perspective of keyword evolution: A case study on joint attention. *Scientometrics* **2018**, *117*, 973–995. [CrossRef]
47. Wang, L.; Notten, A.; Surpatean, A. Interdisciplinarity of nano research fields: A keyword mining approach. *Scientometrics* **2013**, *94*, 877–892. [CrossRef]
48. Xu, H.; Guo, T.; Yue, Z.; Ru, L.; Fang, S. Interdisciplinary topics of information science: A study based on the terms interdisciplinarity index series. *Scientometrics* **2016**, *106*, 583–601. [CrossRef]
49. Kim, K.; Kogler, D.F.; Maliphol, S. Identifying interdisciplinary emergence in the science of science: Combination of network analysis and BERTopic. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 603. [CrossRef]
50. Wang, M.; Xie, Y.; Guo, X.; Fu, H. Interdisciplinarity in the Built Environment: Measurement and Interdisciplinary Topic Identification. *Buildings* **2024**, *14*, 3718. [CrossRef]
51. Chaoguang, H.; Yueji, H.; Fanfan, H.; Chenwei, Z. An approach for interdisciplinary knowledge discovery: Link prediction between topics. *Phys. A Stat. Mech. Its Appl.* **2025**, *665*, 130517. [CrossRef]
52. Omodei, E.; De Domenico, M.; Arenas, A. Evaluating the impact of interdisciplinary research: A multilayer network approach. *Netw. Sci.* **2017**, *5*, 235–246. [CrossRef]
53. Adams, J.; Light, R. Mapping interdisciplinary fields: Efficiencies, gaps and redundancies in HIV/AIDS research. *PLoS ONE* **2014**, *9*, e115092. [CrossRef]
54. Montgomery, D.C.; Jennings, C.L.; Kulahci, M. *Introduction to Time Series Analysis and Forecasting*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
55. Liu, Z.; Zhu, Z.; Gao, J.; Xu, C. Forecast methods for time series data: A survey. *IEEE Access* **2021**, *9*, 91896–91912. [CrossRef]
56. Mudelsee, M. Trend analysis of climate time series: A review of methods. *Earth-Sci. Rev.* **2019**, *190*, 310–322. [CrossRef]
57. Abuhay, T.M.; Nigatie, Y.G.; Kovalchuk, S.V. Towards predicting trend of scientific research topics using topic modeling. *Procedia Comput. Sci.* **2018**, *136*, 304–310. [CrossRef]
58. Xiong, X.; Yang, G.-L.; Zhou, D.-Q.; Wang, Z.-L. How to allocate multi-period research resources? Centralized resource allocation for public universities in China using a parallel DEA-based approach. *Socio-Econ. Plan. Sci.* **2022**, *82*, 101317. [CrossRef]

59. Li, X.; Liu, Y.; Li, A.; Yang, X.; Tang, X. Research on the Subject Mapping Methods Optimization of Multi-Sources Biomedical Literatures. *Digit. Libr. Forum* **2023**, *19*, 1–9.
60. Biglan, A. The characteristics of subject matter in different academic areas. *J. Appl. Psychol.* **1973**, *57*, 195–203. [CrossRef]
61. Ma, L.; Zhao, Y.; Zuo, K.; Yi, P.; Liu, H. Analysis on the development trends of synthetic biology from the interdisciplinary perspective. *Sci. Manag. Res.* **2023**, *41*, 19–26. [CrossRef]
62. Leydesdorff, L. On the normalization and visualization of author co-citation data: Salton's Cosine versus the Jaccard index. *J. Am. Soc. Inf. Sci. Technol.* **2008**, *59*, 77–85. [CrossRef]
63. Gribetz, S.K.; Kaye, L. 3 Time in Disciplinary and Interdisciplinary Perspective. In *Time: A Multidisciplinary Introduction*; Walter de Gruyter GmbH: Berlin, Germany, 2023; pp. 53–82.
64. Frogeri, R.F.; Júnior, P.d.S.P. Interdisciplinary Frontiers: Bridging Knowledge for Contemporary Challenges. *Rev. Myth.* **2025**, *22*, 238–243. [CrossRef]
65. Dalal, A.-A.; AlRassas, A.M.; Al-qaness, M.A.; Cai, Z.; Aseeri, A.O.; Abd Elaziz, M.; Ewees, A.A. TLIA: Time-series forecasting model using long short-term memory integrated with artificial neural networks for volatile energy markets. *Appl. Energy* **2023**, *343*, 121230.
66. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [CrossRef]
67. Dey, R.; Salem, F.M. Gate-variants of gated recurrent unit (GRU) neural networks. In Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, MA, USA, 6–9 August 2017; pp. 1597–1600.
68. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2017; p. 30.
69. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
70. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2021.
71. Yuan, Y.; Meng, Y. Evolutionary Analysis and Frontier Identification of Interdisciplinary Research in Foreign LIS Field Based on BERT-LDA. *Digit. Libr. Forum* **2024**, *20*, 1–15.
72. Dwivedi, Y.K.; Sharma, A.; Rana, N.P.; Giannakis, M.; Goel, P.; Dutot, V. Evolution of artificial intelligence research in Technological Forecasting and Social Change: Research topics, trends, and future directions. *Technol. Forecast. Soc. Change* **2023**, *192*, 122579. [CrossRef]
73. Hain, D.; Jurowetzki, R.; Lee, S.; Zhou, Y. Machine learning and artificial intelligence for science, technology, innovation mapping and forecasting: Review, synthesis, and applications. *Scientometrics* **2023**, *128*, 1465–1472. [CrossRef]
74. Iram, A.; Dong, Y.; Ignea, C. Synthetic biology advances towards a bio-based society in the era of artificial intelligence. *Curr. Opin. Biotechnol.* **2024**, *87*, 103143. [CrossRef]
75. Gong, X.; Zhang, J.; Gan, Q.; Teng, Y.; Hou, J.; Lyu, Y.; Liu, Z.; Wu, Z.; Dai, R.; Zou, Y. Advancing microbial production through artificial intelligence-aided biology. *Biotechnol. Adv.* **2024**, *74*, 108399. [CrossRef]
76. Silva, T.C.; Eppink, M.; Ottens, M. Automation and miniaturization: Enabling tools for fast, high-throughput process development in integrated continuous biomanufacturing. *J. Chem. Technol. Biotechnol.* **2022**, *97*, 2365–2375. [CrossRef]
77. Jiang, W.; Wu, Z.; Gao, Z.; Wan, M.; Zhou, M.; Mao, C.; Shen, J. Artificial cells: Past, present and future. *ACS Nano* **2022**, *16*, 15705–15733. [CrossRef]
78. Dixit, S.; Kumar, A.; Srinivasan, K.; Vincent, P.D.R.; Ramu Krishnan, N. Advancing genome editing with artificial intelligence: Opportunities, challenges, and future directions. *Front. Bioeng. Biotechnol.* **2024**, *11*, 1335901. [CrossRef]
79. Li, Z.; Khan, W.U.; Bai, G.; Dong, C.; Wang, J.; Zhang, Y.; Wang, C.; Zhang, H.; Wang, W.; Luo, M.; et al. From Code to Life: The AI-Driven Revolution in Genome Editing. *Adv. Sci.* **2025**, e17029. [CrossRef]
80. Chen, V.; Yang, M.; Cui, W.; Kim, J.S.; Talwalkar, A.; Ma, J. Applying interpretable machine learning in computational biology—Pitfalls, recommendations and opportunities for new developments. *Nat. Methods* **2024**, *21*, 1454–1461. [CrossRef]
81. Xue, L.; Pang, Z. Ethical governance of artificial intelligence: An integrated analytical framework. *J. Digit. Econ.* **2022**, *1*, 44–52. [CrossRef]
82. Mateos Fernández, R.; Petek, M.; Gerasymenko, I.; Juteršek, M.; Baebler, Š.; Kallam, K.; Moreno Giménez, E.; Gondolf, J.; Nordmann, A.; Gruden, K. Insect pest management in the age of synthetic biology. *Plant Biotechnol. J.* **2022**, *20*, 25–36. [CrossRef]
83. Li, J.; Zhao, H.; Zheng, L.; An, W. Advances in synthetic biology and biosafety governance. *Front. Bioeng. Biotechnol.* **2021**, *9*, 598087. [CrossRef] [PubMed]
84. Sun, T.; Song, J.; Wang, M.; Zhao, C.; Zhang, W. Challenges and recent progress in the governance of biosecurity risks in the era of synthetic biology. *J. Biosaf. Biosecur.* **2022**, *4*, 59–67. [CrossRef]
85. Liao, C.; Xiao, S.; Wang, X. Bench-to-bedside: Translational development landscape of biotechnology in healthcare. *Health Sci. Rev.* **2023**, *7*, 100097. [CrossRef]
86. Brooks, S.M.; Alper, H.S. Applications, challenges, and needs for employing synthetic biology beyond the lab. *Nat. Commun.* **2021**, *12*, 1390. [CrossRef] [PubMed]

87. Hanna, E.; Levic, A. Comparative Analysis of Language Models: Hallucinations in ChatGPT: Prompt Study. 2023. Available online: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1764165&dswid=2283 (accessed on 10 May 2025).
88. Herrera-Poyatos, D.; Zuheros, C.; Montes, R.; Herrera, F. Large language models for crowd decision making based on prompt design strategies using ChatGPT: Models, analysis and challenges. *arXiv* **2024**, arXiv:2403.15587.
89. Kalai, A.T.; Vempala, S.S. Calibrated language models must hallucinate. In Proceedings of the 56th Annual ACM Symposium on Theory of Computing, Vancouver, BC, Canada, June 24–28 2024; pp. 160–171.
90. Evstafev, E. The Paradox of Stochasticity: Limited Creativity and Computational Decoupling in Temperature-Varied LLM Outputs of Structured Fictional Data. *arXiv* **2025**, arXiv:2502.08515.
91. Yu, G.; Guo, R.; Li, Y.-J. The influence of publication delays on three ISI indicators. *Scientometrics* **2006**, *69*, 511–527. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Decoding Wine Narratives with Hierarchical Attention: Classification, Visual Prompts, and Emerging E-Commerce Possibilities

Vlad Diaconita *, Anda Belciu, Alexandra Maria Ioana Corbea and Iuliana Simonca

Department of Economic Informatics and Cybernetics, Bucharest University of Economic Studies, 010374 Bucharest, Romania; anda.belciu@ie.ase.ro (A.B.); alexandra.florea@ie.ase.ro (A.M.I.C.); iuliana.botha@ie.ase.ro (I.S.)
* Correspondence: diaconita.vlad@ie.ase.ro

Abstract

Wine reviews can connect words to flavours; they entwine sensory experiences into vivid stories. This research explores the intersection of artificial intelligence and oenology by using state-of-the-art neural networks to decipher the nuances in wine reviews. For more accurate wine classification and to capture the essence of what matters most to aficionados, we use Hierarchical Attention Networks enhanced with pre-trained embeddings. We also propose an approach to create captivating marketing images using advanced text-to-image generation models, mining a large review corpus for the most important descriptive terms and thus linking textual tasting notes to automatically generated imagery. Compared to more conventional models, our results show that hierarchical attention processes fused with rich linguistic embeddings better reflect the complexities of wine language. In addition to improving the accuracy of wine classification, this method provides consumers with immersive experiences by turning sensory descriptors into striking visual stories. Ultimately, our research helps modernise wine marketing and consumer engagement by merging deep learning with sensory analytics, proving how technology-driven solutions can amplify storytelling and shopping experiences in the digital marketplace.

Keywords: wine classification; hierarchical attention networks; GloVe embeddings; sensory descriptors; AI-driven marketing

1. Introduction

In the digital age, unstructured text data have grown rapidly, necessitating the use of advanced processing and analysis techniques. Topic labelling, the process of assigning various unstructured text labels that correlate with subjects or categories, is a significant application of text categorisation. Its main goal is to extract general tags in order to facilitate a variety of analytical applications. In addition, there are numerous possible applications for this strategy in numerous fields, each of which has its own distinct set of opportunities and challenges.

Content tagging apps use these techniques to improve user experience and search engine rankings by automatically classifying a range of digital content. Additionally, text classification aids in common applications such as automated response generation and email spam detection. By grouping user content, businesses can identify emerging trends and more accurately characterise different types of customers. Specifically, by enabling the extraction of valuable information from customer comments, text classification removes

the need for a time-consuming human evaluation in the product review process. Sentiment analysis, a related field, goes even further by predicting people's feelings about specific topics based on text input. This is helpful for social and political discourse analysis in addition to economic analysis. For example, Ferrara and Ciano (2024) [1] recently demonstrated that short-term ESG stock price forecasts are considerably sharpened when explainable AI models are combined with news and social media signals, which serve as a proxy for market sentiment.

Storytelling skills are becoming more and more important in digital wine retail. Price, availability, and consumer trust can all be impacted by a single expert review; however, these reviews are often long, full of metaphors, and written in a vocabulary (tannin, terroir, finish, etc.) that is difficult for sentiment analysis tools to understand. Therefore, it is both a linguistic and a business challenge to transform such language into organised, e-commerce-ready information. In relation to this, Croijmans et al. [2] performed a computational analysis of reviews by wine experts and discovered that despite the assertions that wine descriptions lack information, experts consistently and informatively describe wines. Their research demonstrated how expertise can overcome limitations in describing sensory experiences like flavour and smell by revealing domain-specific vocabulary and the capacity to predict wine properties from review text. Reviews serve as dynamic quality signals in markets, as evidenced by the strong correlation between consumer price responses and expert wine ratings [3].

The main goal of this study is to investigate the use of state-of-the-art neural network architectures for the classification of wine reviews and the extraction of key attributes that characterise the quality and style of wine. By making it easier to create eye-catching marketing ads and enhance websites, these techniques seek to support numerous facets of data analysis and decision-making processes, as well as the marketing process. The usefulness of deep learning methods in sentiment analysis, especially for social media content, has been reaffirmed by recent studies. For example, a comprehensive study comparing different deep learning architectures embedded with Word2Vec and GloVe techniques was carried out by Rakshit and Sarkar [4]. Their results demonstrate how well Word2Vec-embedded models, particularly bidirectional recurrent architectures such as Bi-LSTM and Bi-GRU, classify tweet polarity. Furthermore, recent research demonstrates that graph convolutional network methods for text classification perform well on domain-specific tasks [5]. Similarly, Prakash and Vijay [6] show that hierarchical transformer-based sentiment analysis frameworks can combine explainability and aspect extraction, resulting in precise predictions and clear identification of unique sentiment signals for a given domain. These studies support the potential of cutting-edge machine learning techniques to enhance text classification and sentiment analysis skills across a range of applications. Other research [7], suggesting a natural connection between textual review extraction and the automated production of e-commerce marketing assets, shows that generative AI systems can produce more persuasive visual content than human marketers. However, conventional bag-of-words or flat RNN classifiers frequently misinterpret domain-specific polarity ('dry' is not a negative term) and fail to recognise the hierarchical structure of wine notes (word $\rightarrow$ sentence $\rightarrow$ review). As a result, prior research either avoids the issue of how text insights should be incorporated into marketing materials or hand-codes small lexica.

We present an end-to-end pipeline that (i) uses a Hierarchical Attention Network to classify expert reviews, (ii) uses a domain-specific lexicon to recalibrate sentiment, and (iii) transforms the resulting descriptors into AI-generated visuals that are evaluated in order to fill these gaps. The objective is to demonstrate how narrative language can be recoloured into content aimed at consumers after being grayscaled in the data. Four

interrelated research questions (RQs) that each focused on a crucial component of the pipeline's architecture, functionality, and use were used to examine this.

RQ1 How is the perceived stylistic difference between the two image sets related to (a) participants' overall set-level preference and (b) the gap in their mean relevance/appeal ratings?

RQ2 Which self-reported rating factors are predictive of appeal and relevance scores, and do these predictive weights change depending on wine marketing experience?

RQ3 In terms of categorising expert wine reviews by style and quality, how does our Hierarchical Attention Network (with and without GloVe augmentation) stack up against the Spectral-HAN and LSTM baselines?

RQ4 When compared to a generic sentiment tool, does recalibrating sentiment scores using a wine-tuned lexicon improve aspect-level accuracy and decrease polarity errors?

Our primary contributions, which connect linguistic analysis, model innovation, and real-world e-commerce applications, are guided by these RQs:

(1) A web-based user study assessing stylistic preferences and rating factors validates the efficacy of our prompt-engineering pipeline for RQ1 and RQ2, which demonstrates a proof-of-concept pipeline that converts key descriptors into AI-generated images.

(2) We develop a hierarchical attention classifier that surpasses flat LSTM and Spectral-HAN baselines and also rivals published DistilBERT results (addressing RQ3).

(3) We release a cleaned corpus of 46,000 Wine Enthusiast reviews, plus the full web interface and assets used in our user study, to enable reproducible benchmarking and domain-specific sentiment research (supporting RQ3 and RQ4).

The state of the art in sentiment analysis, wine informatics, and text classification is reviewed in Section 2. The dataset, neural architectures, and evaluation methodology are explained in Section 3. The experimental results are presented in Section 4, and then the commercial implications, limitations, and future work directions are examined in Section 5. The study is finally concluded in Section 6.

2. Literature Review

Natural language processing (NLP) and machine learning have fueled recent advances in text classification and sentiment analysis, enabling the extraction of useful details from large amounts of unstructured data. This section examines the evolution of text categorisation methods, explores the creation of neural network designs, and illustrates how these tools are applied to the analysis of complex textual domains such as wine reviews.

2.1. Evolution of Text Classification and Topic Labelling

Text classification, the process of assigning predefined categories to textual data, has been a foundational task in natural language processing (NLP). Many approaches rely on statistical and machine learning methods, such as Random Forest, Naïve Bayes classifiers and Support Vector Machines (SVMs) for tasks such as spam filtering [8], content filtering [9], or document classification [10]. One interesting advancement in the area was the Latent Dirichlet Allocation (LDA), which represents documents as mixtures of topics, where each topic is characterised by a distribution over words [11]. It works especially well for uncovering hidden patterns in large datasets, making it possible to find recurring themes in documents [12]. By adding more statistical data for better topic coherence, LDA variants, like the Doubly Correlated Topic Model, expand on their potential [13]. Deep learning and clustering methods are frequently used for topic modelling [14].

It is necessary to understand the challenges related to class imbalance or label ambiguity, including for the evaluation of results. Some studies investigated the relationship

between changes in a classifier's confusion matrix and the resulting metrics. One such study introduces the concept of "measure invariance," a critical property that determines whether a performance measure is sensitive to certain changes in classification outcomes [15]. Tasks involving human communication, such as sentiment analysis, may require measures, like the area under the curve (AUC), that are more sensitive to shifts in both positive and negative classifications.

2.2. Sentiment Analysis and Its Diverse Applications

Sentiment analysis involves determining the emotional tone behind a body of text, ranging from positive and negative sentiments to more nuanced emotional states. It has applications across domains, including market research, social media monitoring, and customer feedback analysis [16]. Traditional sentiment analysis methods employed lexicon-based approaches, which matched words with predefined sentiment scores. However, these methods often failed to account for context, sarcasm, and idiomatic expressions.

Deep learning models have significantly advanced sentiment analysis by learning from large datasets to recognise complex patterns and contextual cues. The effectiveness of embedding techniques like Word2Vec and GloVe combined with bidirectional architectures such as Bi-LSTM and Bi-GRU in capturing the polarity of tweets has been demonstrated [4]. This highlights the importance of contextual embeddings in improving sentiment classification accuracy.

Aspect-based Sentiment Analysis (AbSA) has emerged as a specialised domain of sentiment analysis, offering a more granular approach by analyzing sentiments at the level of specific aspects or entities within text. Unlike traditional sentiment analysis, which often evaluates the overall tone of a document or sentence, AbSA focuses on determining sentiments tied to distinct aspects, enabling a better understanding of user opinions [16–18]. Despite its potential, the field presents several challenges that continue to inspire research and innovation.

One major challenge in AbSA is aspect extraction (AE), which involves identifying explicit aspects, implicit aspects, and Opinion Target Expressions (OTEs). Researchers have explored various machine learning and deep learning approaches to improve AE. However, handling cross-domain and cross-lingual scenarios remains a significant hurdle, and effective mapping of relationships between aspects and sentiments is an ongoing concern [19].

Similarly, the classification process, which assigns sentiment polarities to extracted aspects, faces its own set of difficulties. Models must account for interactions, dependencies, and contextual–semantic relationships between aspects, sentiment words, and surrounding text. Challenges such as multiword targets and managing inter-aspect dependencies remain active areas of research [20].

Another emerging area within AbSA is sentiment evolution (SE), which seeks to understand and predict how sentiments about aspects change over time using mechanisms such as social network analysis or epistemic network analysis. Recognising dynamic factors such as cognitive attitudes, social behaviours, and external influences poses a significant challenge. Additionally, predicting these changes in dynamic social media environments or within a blended learning context is an open problem that requires further exploration [21].

2.3. Advanced Neural Network Architectures for Text Analysis

Many deep learning architectures have facilitated text analysis innovations in recent years. Originally designed for image recognition, Convolutional Neural Networks (CNNs) have been reinterpreted for NLP applications by means of convolutional filters over embedded matrices to gather local textual information. Especially in large annotated corpora,

these filters greatly improve classification performance and are effective at spotting n-gram-like patterns [22]. CNNs by themselves, nevertheless, might occasionally find it difficult to handle longer dependencies inherent in natural language.

Using gated techniques to store and retrieve information across long text sequences, Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks [23] and their bidirectional variants (i.e., BiLSTM), address this problem. Effective in representing temporal and contextual interactions, these designs fit text categorisation, sentiment analysis, and other applications using sequential data. Nevertheless, because of vanishing or exploding gradients, conventional RNNs can still find difficulty processing extremely long documents.

Most famously BERT transformer-based models enhance these powers through multi-head attention systems, enabling them to manage whole sequences in concurrently rather than step by step. BERT achieves state-of-the-art performance on several NLP benchmarks by learning context from both left and right tokens using a bidirectional training target. Transformers can, however, sometimes overfit smaller domain-specific datasets and can be computationally costly; hence, careful fine-tuning is advised.

Hierarchical Attention Networks (HANs) [24] mix recurrent units with attention at both word and sentence levels. HANs give better-grained control over which phrases or sentences are most important in classification by modelling text in a hierarchical manner, with words arranged into sentences and then sentences arranged into documents. For wine reviews, which frequently feature several lines with different descriptions of flavour, scent, and quality, this is particularly important. Moreover, attention mechanisms enable the model to focus especially on domain-specific vocabulary, thus possibly catching minor variations in the wine language [25].

These cutting-edge designs, CNNs, RNNs (LSTM/BiLSTM), transformers (BERT), and hierarchical models (HAN) offer, all around, complementary methods to text analysis. Their interaction with rich word embeddings (e.g., Word2Vec, GloVe) improves semantic understanding, therefore facilitating more accurate classification of complex material, including expert wine evaluations. The next sections address how these designs might be modified and assessed for the specific goal of wine review classification.

2.4. Analyzing Wine Reviews: Challenges and Techniques

2.4.1. The Complexity of Wine Language

When it comes to analysing wine categories based on evaluations, one must first understand the wine language, or winespeak. People who often drink or taste wine have their own vocabulary when referring to the acidity of the wine, sweetness, level of alcohol, flavour, colour, or other wine properties. Of course, its flavour or taste is the most subjective and generous characteristic, having associations with fruits, flowers, spices, herbs, or chocolate. This feature is quite similar with coffee picking. They both have resemblances in the arts where music, paintings, or sculptures may have different interpretations due to the subjectivity, sensibility, or experience of the consumer.

Reviews of the literature on wine reviews are performed in different papers to determine if the language and expressions used in such reviews influence the price of the wines. For instance, Ref. [26] tests the following hypotheses supported by the study: Higher-priced wine reviews will be more verbose; wine reviewers will use more emotional language with lower-priced wines; wine reviewers will use more logical language with higher priced wines; and wines of higher price will have more social language in reviews. It was stated that sophisticated reviews usually appear for expensive and therefore better wines, while concise and straightforward comments accompany cheap wines. In any case,

understanding wine language increases a consumer's confidence in selecting a wine kind and knowing what to expect.

Online marketing efforts became extremely important and necessary as the COVID epidemic raised the user count of social media networks and the time spent online. Any product can become prominent and viral with great promotion and sharing. The social internet caused surprising points of view or products to emerge suddenly. One study [27] revealed how wine companies use Instagram in response to visual communication tools and demand to be in trend with brand reputation showing. It underlines the need of not just the language one uses but also of who speaks it and over what channels of communication. Another study [28] examined the digital presence and consumer engagement of 311 Greek wineries on Facebook and Instagram between 2019 and 2022, including during the COVID-19 epidemic. The results show consistently low degrees of interaction on wineries' profiles, which emphasises the need for more study on winery marketing strategies and implies that the suggested three-phase social media analytics methodology can act as a model for improving brand engagement across many sectors.

Red and white wines were lexically analysed from the reviews points of view [29]. Wine expert reviews are more analytical and with many details when it comes to red wines and more impersonal and credible for white wines. Therefore, white wine reviews are considered more reliable.

Whether the reviewer is expert or not, they might use many metaphors when describing the properties of wine, depending on their knowledge and background. Expressions coming from textiles (e.g., silky, velvety, satiny, pillowy), architecture (e.g., structure and balance), physiology (e.g., breeding or ageing), and anatomy (e.g., skeleton or spine) may occur in the language that describes wine [30].

A quantitative analysis indicated that wine reviewers mainly use the WAW (Wine Aroma Wheel) and WSET (Wine and Spirit Education Trust) standard vocabularies composed by oenologists in 1984 and 2023 [31]. Only 10.5% of reviewers have other wine language patterns, which include the terms "juicy", "dark", "crisp", "firm", and "structured" with the highest frequency.

2.4.2. Machine Learning and Neural Networks in Wineinformatics

At the intersection of data science, artificial intelligence, and oenology, the field of wineinformatics was proposed. It is used mostly for wine quality prediction, but also it has other applications like market analysis, viticulture optimisations, or flavour profiling. Wine quality prediction is performed using machine learning models on reviews, feedback, or wine competitions.

Machine learning applications in oenology include quality prediction, variety classification, market analysis, and viticulture optimisation. Comparative studies demonstrate the effectiveness of ensemble methods in wine quality prediction. According to one study, AdaBoost and Random Forest excel in classifying Pinot Noir [32], while another finds Random Forest to be especially successful at predicting white wine quality [33].

Classifying numerical data is relatively straightforward, but the textual elements of wine reviews and feedback introduce linguistic nuances that demand more sophisticated methods. Neural networks, including CNN, BiLSTM, and BERT, have proven particularly adept at modelling these complexities. For instance, research applying ANNs to wine technology illustrates their ability to uncover intricate patterns in data that simpler approaches may miss [34].

Text analysis in wine research has shifted from traditional machine learning to deep learning, which employs architectures such as Feedforward Neural Networks (FNNs), Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), Generative

Adversarial Networks (GANs), and transformer networks (such as BERT). One study [35] used acidity, sugar, and alcohol content to predict red and white wine quality. Another dual-method technique employed exploratory data analysis and a VGG19 CNN [36] to estimate the likelihood of distinct quality groups, demonstrating deep learning's ability to capture complex connections between chemical properties and overall quality. NLP research has also showed potential in the analysis of consumer-facing wine data such as ratings, reviews, and social media posts. A recent study [37] developed a recommendation system that uses BERT to identify wine varietals, hence raising consumer expectations and overall quality perception.

Data from an electronic nose (i.e., an intelligent sensor system utilised in the food and beverage industries) was examined [38] utilising Artificial Neural Networks (ANNs) and basic machine learning techniques such as KNN and SVM. A similar detection method employed a backpropagation neural network (a form of FNN) to classify wines and liquors [39]. Deep learning models have demonstrated remarkable effectiveness in extracting knowledge from tough datasets, frequently improving through synergy with other analytical methods. For example, merging feedforward neural networks with evolutionary algorithms resulted in accurate prediction of fermentation factors [40]. Another study [41] used molecular composition and ANN-based spectroscopic analysis to categorise grape varietals and measure ripeness, while research on hyperspectral images of grape fruit [42] shows that CNNs, Inception Time Networks, and MiniRocket can estimate sugar content.

In terms of market analysis, text mining wine reviews indicated diverse customer profiles across countries. The American market loves berry-like, chocolate, and citrus overtones, whilst the British market prefers jammy, berry, cherry, and dark chocolate flavours. Chinese consumers want refreshing, delicate flavours with mellow scents [43]. Consumer reviews have also been harvested from Twitter to identify sentiment towards red wines [44]. In terms of variety categorisation, Lefever et al. [45] employed text mining on wine reviews to forecast a wine's colour, grape variety, and country of origin. Chemometric and sensometric approaches [46] and language processing techniques [47] have also been used.

Recent advances in wineinformatics show a growing trend toward integrating sophisticated data analysis techniques into oenology. One notable approach employs a tri-clustering algorithm that discovers latent groupings spanning wine characteristics, attributes, and vintage information [48]. Furthermore, other researchers used Naïve Bayes and SVM classifiers to unravel the complexities inherent in Bordeaux wines [49]. Adding further depth to the field, a recent contribution introduces the Computational Wine Wheel, a natural language processing tool specifically designed to sift through wine reviews [50]. However, neural networks outperformed both Naïve Bayes and SVM on a Bordeaux dataset comprising more than 10,000 wines [51].

Most wine reviews focus on quality predictions or on classical text mining using AdaBoost, Random Forest, Siamese Neural Networks, or CNNs. Hierarchical Attention Networks (HANs) were poorly applied to wine reviews, and thus using HAN models for wine classification reflects the complexities of the wine language.

Extrapolating, classifying wine reviews can also be applied to assortments of coffee, chocolate, beer, or tea and even to food pairing (using Generative Adversarial Networks for approaching generative artificial intelligence [52] or Siamese Neural Networks, i.e., twin neural networks [53]).

Collectively, these studies highlight the promise of deep learning techniques for identifying links and trend analysis in wine datasets. They yield scalable, accurate solutions for quality assessment and control while also proving that wine language can predict key features, with review length and word choice influencing pricing and rating.

2.5. Bridging Textual Analysis and Visual Content Generation

2.5.1. Generative Models and Marketing Innovation

The field of digital marketing is one of the most dynamic areas of growth for a company and recent developments in technology coupled with the widespread accessibility of digital devices are leading to its evolution from a subcategory of classic marketing to a new, almost independent field. As V. Soni [54] highlights, digital marketing is increasingly focusing on the user, and measurable interactive and generative models are playing an essential role in this evolution.

Generative artificial intelligence models (GenAI) offer revolutionary capabilities for marketing, a field that, according to [55], is predicted to experience the greatest impact from this technology, with projected productivity gains of up to 15% in marketing expenditure, translating to an estimated $463 billion annually. GenAI can generate personalised marketing content, such as product descriptions, emails, and advertisements tailored to individual preferences and behaviours [56], it can produce dynamic advertisements that adapt in real time based on user behaviour or market trends, or it can automate the analysis of consumer feedback, social media data, and market trends [57]. Thus, it is surprising that many companies have already begun leveraging GenAI to carry out essential and innovative marketing initiatives. A study by [58] that involved several marketing professionals showed that all the specialists interviewed had already explored the possibilities offered by different text-to-image models, such as DALL-E, Midjourney, or Stable Diffusion.

The DALL-E model can convert textual inputs into vivid visual representations. By mapping abstract textual descriptions into detailed images using embeddings derived from language models, such as GloVe or GPT, models like this bridge the gap between linguistic data and creative visual outputs, thus facilitating the creation of targeted advertisements and enabling deeper consumer engagement. Studies show that personalised marketing materials generated using AI increase click-through and conversion rates by more than 30% [59]. In addition, using AI-generated content not only reduces production costs and time but also increases efficiency, empowering companies to innovate, foster engagement, and expand their reach in a competitive market.

2.5.2. Applications in the Wine Industry

The wine industry, known for its deep connection to tradition and craftsmanship, would greatly benefit from exploring the potential of GenAI to implement innovation in areas such as marketing, product design, and consumer education while maintaining their artisanal roots.

We have already established that wine aficionados have their own rich vocabulary when describing different product characteristics and this vocabulary usually references associations with vivid imagery. Wine reviews are often times almost artistic in their description of the product and the wine industry can benefit from combining textual analysis of reviews with visual content generation. By extracting key descriptors from reviews, wineries can create imagery that reflects the sensory attributes and narratives associated with their products.

By extracting key descriptors from reviews, wineries can create imagery that reflects the sensory attributes and narratives associated with their products. As Refs. [60,61] emphasise, the connection between sensory descriptors and visual narratives fosters a stronger emotional bond with consumers. Using generative models like DALL-E [62] to convert sensory descriptors into visually compelling imagery would enhance brand perception and consumer loyalty. The consumers' understanding and appreciation of complex wine characteristics can be significantly enhanced by translating sensory data into visuals. For example, a wine described as "crisp with notes of green apple" can be represented with imagery that reflects freshness and lightness [63].

Another side of the wine industry in which generative models could have a significant impact is the design of wine labels, which represent a critical marketing tool allowing products to stand out in competitive markets. The best use of AI capabilities in this area would be to generate customised label designs based on sensory keywords extracted from wine reviews. Ref. [64] indicated that aesthetically engaging labels significantly influence consumer purchasing decisions, making this application both practical and profitable. Moreover, when creating such personalised designs, specific cultural and regional preferences, both general and specific ones resulting from textual product reviews, can be more easily considered. For example, in the Chinese market, where elegance and tradition are attributes highly appreciated by consumers, generative models could create designs that include subtle patterns and traditional motifs that align with local tastes [43]. By bridging cultural nuances with advanced AI technologies, wineries can expand their global appeal.

2.6. Applications of Neural Networks in E-Commerce

There is little doubt that e-commerce's future is data driven, customer-centric, and highly experiential. AI models enable marketers to listen to the customer's voice at scale, distilling actionable sentiment and intent signals from a constant flow of data.

Recent reviews on machine learning (ML) and deep learning (DL) in e-commerce show several research directions, including sentiment analysis, recommendation systems, disinformation detection, fraud detection, prediction of customer churn, and product classification [65–67]. Researchers increasingly employ transformer-based models, such as BERT and GPT, that have the ability to capture nuanced contextual meanings, outperforming earlier approaches based on RNNs or CNNs. Hierarchical Attention Networks (HANs) can effectively summarise lengthy customer reviews by identifying critical sentences or aspects [68]. Deep learning techniques and neural factorisation can aid in personalisation, enabling e-commerce platforms to offer more precise product recommendations adapted to individual user behaviour [69]. Meanwhile, CNN-based and hybrid neural network approaches can distinguish genuine reviews from manipulated reviews, reinforcing trustworthiness on e-commerce platforms [70] with the help of rules-based classifiers. Although powerful, these models present ongoing challenges, including multilingual application difficulties, interpretability issues, and computational complexity, particularly with respect to real-time processing [71].

Another recent trend in digital commerce is AI-driven marketing personalisation. Researchers have been developing sophisticated recommender systems and targeting algorithms that go beyond one-size-fits-all approaches. Yıldız et al. (2023) [72] present a hyper-personalised product recommendation system for fashion retail. Their approach integrated multiple data mining techniques: (1) RFM-based customer segmentation to group shoppers by behaviour, (2) association rule mining on each segment (considering location and purchase history) to learn preferences, and (3) a recommendation engine tailored to each segment's profile. By segmenting customers and then providing group-specific recommendations, this system attempts to address the gap between individual personalisation and scalable marketing. The result is a more context-aware recommender that can, for example, suggest seasonal apparel relevant to a customer's region and buying patterns. In this regard, the common thread is using AI to learn from consumer interactions (clicks, views, past purchases) and then optimise the marketing mix for each consumer.

2.7. Synthesis and Research Gap

As established through the literature review, significant progress has been made in textual and sentiment analysis, enhancing their applicability across diverse domains, including wine reviews. These advances allow us to better understand subjective content,

with new subdomains such as AbSA (Aspect-based Sentiment Analysis) giving us the opportunity to capture finer details regarding the consumer perceptions of wines, wineries, or producers.

Likewise, developments in neural network topologies, notably Convolutional Neural Networks, Inception Recurrent Convolutional Neural Networks, and BERT models, have pushed the boundaries of text analysis capabilities providing reliable and accurate predictions and demonstrating robust generalisation abilities.

However, there are also a few challenges that must be considered gaps when looking at the application of such techniques, especially in the particularly complex field of oenology.

Research in computational wine analysis has revealed significant challenges in processing review text. Current methods struggle with context-dependent interpretations and regional variations in wine descriptions. These linguistic complexities particularly affect Aspect-based Sentiment Analysis, as overlapping references and subjective expressions make it difficult to accurately identify and classify wine components. Reviewer preferences and cultural backgrounds further compound these challenges, affecting how sentiments are expressed and interpreted across different markets.

Despite these advances, gaps exist in the application of advanced text categorisation and sentiment analysis techniques to specialised domains such as wine reviews. Current models can struggle to capture the subtle vocabulary, complicated descriptions, and industry-specific terminology found in wine-related literature. Aspect-based Sentiment Analysis (AbSA) can have difficulties in accurately extracting and understanding implicit aspects and multi-word expressions specific to this discipline. Furthermore, the issue of class imbalance exists in wine evaluations, where specific sentiment classes or characteristics may be underrepresented, compromising the strength of classification algorithms. The distinct vocabulary and interpretive elements of wine reviews pose unique analytical challenges. More advanced language models would enhance our understanding of wine evaluation and description practices.

Deep learning models are widely employed in text analysis; however, the adoption of advanced models like Hierarchical Attention Networks (HANs) for wine review classification has been relatively limited. In instances where the HAN has been applied, its effectiveness in sensory descriptor extraction has varied, suggesting a need for more tailored approaches.

Another related area for which relatively limited research has been conducted so far is the one regarding the possibilities of connecting textual analysis and visual content generation. Generative AI has shown considerable promise in developing visual content for personalised marketing materials. Furthermore, several advantages of using generative models such as DALL-E or Midjourney have been clearly identified, such as increasing conversion rates, reducing necessary production resources, and increasing efficiency. However, there is still a scarcity of literature on combining textual analysis with other data modalities, such as sensory profiles or regional characteristics, to enhance wine categorisation.

Moreover, the scarcity of large-scale, annotated datasets specific to wine reviews presents a significant barrier to the development and validation of robust text analysis models.

The goal of our research is to further explore the application of advanced neural network architectures to classify wine reviews and extract the salient characteristics that define the quality and style of wine, as well as to investigate the possibilities of automatically generating narratives based on topic modelling results. To achieve this, we scraped a large dataset from Winemag.com, which includes over 46,000 wine reviews collected between 2020 and 2023. This dataset contains diverse attributes such as ratings, descriptions, prices, and wine origins, providing a rich foundation for our analysis. We will evaluate the performance of four different neural network models on a wine classification task:

the Hierarchical Attention Network (HAN), the Spectral Hierarchical Attention Network (SpectralHAN), HAN augmented with pre-trained GloVe embeddings (HAN + GloVe), and a Long Short-Term Memory network with GloVe embeddings (LSTM + GloVe).

3. Methods

Figure 1 presents the general methodology of our study, illustrating the process from data collection through analysis to the generation of marketing content. The process consists of five main stages: (1) data collection and pre-processing of wine reviews, (2) sentiment analysis using multiple tools, (3) neural network modelling with four different architectures, (4) characteristic extraction from high-quality reviews, and (5) marketing content generation. Each stage builds upon the results of previous stages, creating a framework for wine review analysis and marketing content generation. The following subsections detail each component of this pipeline.

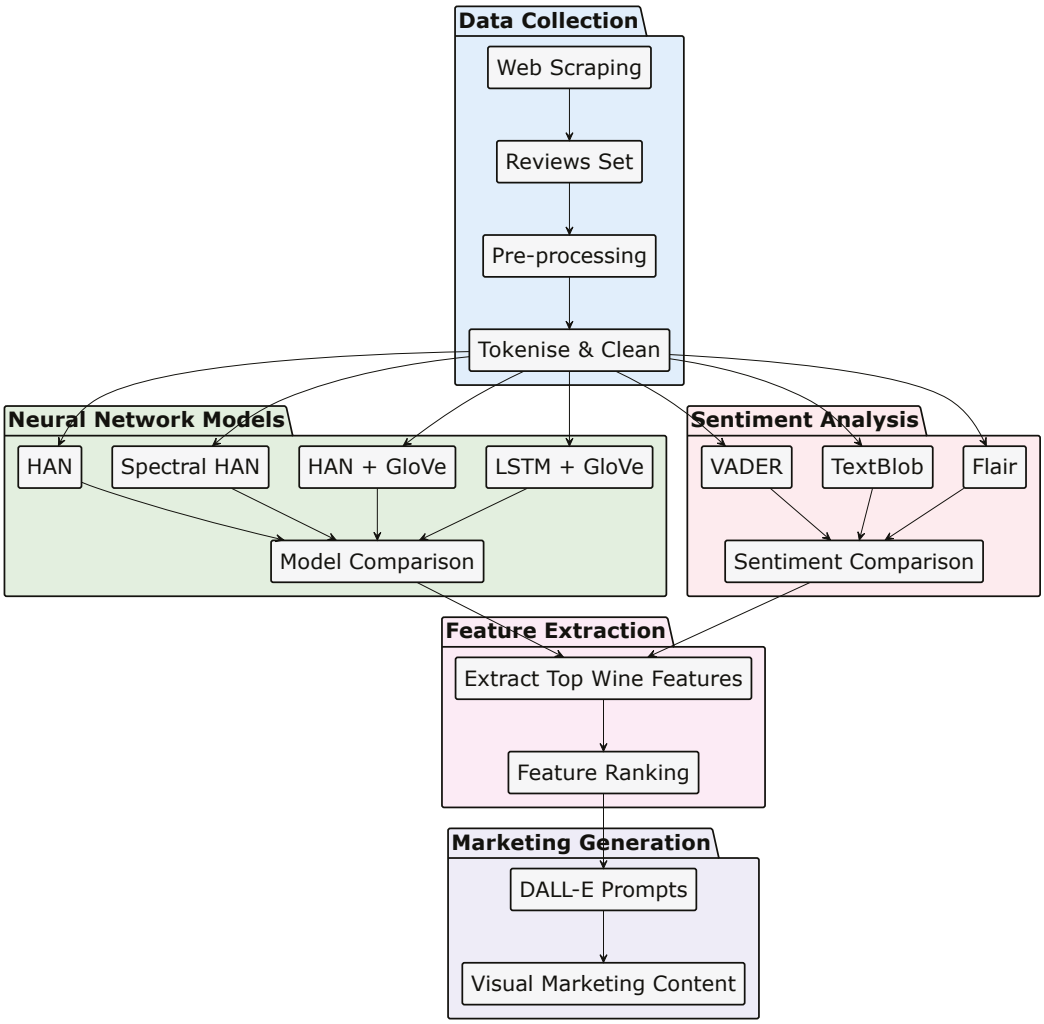


Figure 1. The overview of the methodology.

3.1. Data Scraping

We constructed a web scraper in Python 3.9.12 to collect wine reviews from Winemag.com. It uses BeautifulSoup to parse HTML content from multiple pages, extracting details like ratings, descriptions, prices, and wine origins. The structure of the extracted data is shown in Table 1.

Table 1. The structure of the collected data.

Column Name	Data Type	Example
Points (the rating)	int64	92
title	string	Eberle 2021 Eberle Estate Vineyard Chardonnay (. . .)
Description (the review)	string	This is a solid Chardonn...
taster_name	string	Matt Kettmann
taster_twitter_handle	string	@mattkettmann
taster_photo	string	https://www.winemag.com/wp-content/uploads/2019. . .
Price (in USD)	float64	30.0
designation	string	Eberle Estate Vineyard
variety	string	Chardonnay
region_1	string	Paso Robles
region_2	string	Central Coast
province	string	California
country	string	US
winery	string	Eberle
year	int64	2022

Table 2 presents the number of wine reviews scraped over a four-year period from 2020 to 2023.

Table 2. Number of reviews by year.

Year	Number of Reviews
2020	15,648
2021	19,139
2022	9559
2023	2230
Total	46,576

The country distribution of the reviewed wine is shown in Table 3.

Table 3. The country distribution of the winery's country.

Winery's Country	Number of Reviews
US	21,648
France	10,089
Italy	6103
Portugal	2813
Spain	1525
Australia	1094
Austria	785
New Zealand	578
Chile	490
Argentina	480
South Africa	266
Israel	186
Uruguay	107
Romania	81
Other countries	331

As highlighted in Table 3, in our dataset, the United States has the highest number of winery reviews, with more than 21k, significantly exceeding France and Italy. Other countries like Portugal, Spain, and Australia also have notable representation, while smaller wine-producing countries such as Uruguay and Romania are also represented.

After the scraping process was completed, the data was saved to a JSON file which was used for the next steps of the methodology.

3.2. Data Preprocessing

The data were read from the JSON file constructed during the data scraping step. By using NLTK, a popular text processing Python library, we initiated our preprocessing by tokenising text and applying fundamental cleaning steps—converting to lowercase and removing common words (like "the", "is", "at") and punctuation marks. This created a cleaner corpus where each review was represented as a sequence of meaningful words.

To build our vocabulary, we analyzed how frequently each word appeared across the training dataset. Words appearing less than five times were filtered out to reduce noise and keep the vocabulary manageable. We then created a system to convert words into numbers that our model could understand, including special markers to padding shorter texts and handling new words not seen during training.

Given the hierarchical nature of our model architecture, we used a two-level structure to process reviews. Each review was broken down into sentences, with a limit of 15 sentences per review. We limited the word count in each sentence to 50. Reviews that exceeded these limits were trimmed, while shorter ones were padded to ensure consistent dimensions. This hierarchical structuring balances structural information with fixed input sizes via padding and truncation.

3.3. Sentiment Analysis

VADER, TextBlob, and Flair are tools for sentiment analysis, each with its own approach, strengths, and weaknesses, as further discussed in Table 4. VADER [73] and TextBlob [74] are mainly lexicon-based tools for sentiment analysis. VADER evaluates the emotional intensity of words and phrases, providing sentiment scores on a scale from -1 (most negative) to 1 (most positive), with 0 indicating neutral sentiment. TextBlob offers comparable sentiment analysis but also supports aspect-based analysis, allowing it to identify specific subjects or topics in the text and assess their associated sentiment polarity. Flair [75] employs neural network models and contextual string embeddings, allowing it to capture more nuanced sentiment based on the surrounding context of words and phrases. These techniques have been widely used together or apart to capture the public sentiment on a wide range of problems from sustainability [76] to political discourse [77].

Table 4. Comparison between the sentiment analyses tools.

Aspect	**VADER (Lexicon + Rules)**	**TextBlob (Lexicon/Hybrid)**	**Flair (Deep Learning)**
Approach	Lexicon with rule-based heuristics (punctuation, capitalisation, negation) [73]. No training required; optimised for social media [78].	Pattern-based lexicon assigning and averaging word polarities. Includes subjectivity scoring and optional Naïve Bayes classifier [79].	Neural network (character-level LSTM) with contextual embeddings. Pre-trained on IMDB corpus; supports domain-specific fine-tuning [80].
Strengths	Excels on informal text; handles emoticons and emphasis. Captures sentiment intensity variations. Quick deployment with interpretable outputs. High accuracy on social media content.	User-friendly API for rapid prototyping [81]. Provides polarity and subjectivity metrics. Effective for formal, well-written content. Lightweight with extensible lexicon.	Context-aware with advanced pattern recognition [82]. Superior accuracy on nuanced/lengthy text. Transfer learning capabilities. Multi-language support via embeddings.

Table 4. *Cont.*

Aspect	VADER (Lexicon + Rules)	TextBlob (Lexicon/Hybrid)	Flair (Deep Learning)
Weaknesses	Limited contextual understanding. Primarily English-focused. Struggles with complex negation and irony. Simplistic handling of mixed sentiments.	Poor performance with informal language. Limited contextual analysis. Lower accuracy than VADER on social media. Primarily English-oriented.	Computationally intensive. Binary classification by default. Domain-sensitive performance. Black-box interpretability issues.
Efficiency	High efficiency with minimal resources. Processes thousands of texts per second on CPU.	Fast dictionary-based lookups. Similar performance to VADER.	Significantly slower; requires substantial memory. Benefits from GPU acceleration.
Applications	Social media monitoring, customer reviews, informal content analysis, training data generation.	Survey analysis, educational feedback, labeling, news sentiment tracking, baseline benchmarking [83].	Legal documents, literary analysis, affective computing, multilingual applications, research implementations.
Recent Trends	Multilingual adaptations [78], integration with ML systems, continuing relevance for social content analysis.	Stable algorithm with ongoing use as research benchmark. Some community multilingual extensions. Still updated for compatibility with new Python/NLTK versions (maintenance updates).	Integration with transformers, expanded language support, efficiency improvements, competitive positioning against newer models, including LLMs [79].

Using the tools from Table 4 we assessed the sentiment polarity of the reviews from our dataset and analyzed the results through correlation matrices, descriptive statistics, and percentages of positive sentiments identified by each tool. The results are shown in Table 5.

Table 5. Statistics of sentiment analysis.

Metric	Points	Price	VADER	TextBlob	Flair
Correlation Matrix					
Points	1.000	0.324	0.316	0.131	0.282
Price	0.324	1.000	0.069	0.003	0.048
VADER	0.316	0.069	1.000	0.510	0.365
TextBlob	0.131	0.003	0.510	1.000	0.232
Flair	0.282	0.048	0.365	0.232	1.000
Descriptive Statistics					
Count	-	-	46,572	46,572	46,572
Mean	-	-	0.500	0.151	0.820
Std Dev	-	-	0.368	0.171	0.547
Min	-	-	−0.939	−1.000	−1.000
25%	-	-	0.255	0.033	0.999
Median	-	-	0.599	0.148	1.000
75%	-	-	0.807	0.261	1.000
Max	-	-	0.993	1.000	1.000
% Positive	-	-	80.94%	80.21%	91.10%

The correlation between sentiment scores and critic ratings is low, 0.316 for VADER, 0.131 for TextBlob, and 0.282 for Flair, showing that a high sentiment score does not necessarily map to a high wine rating and vice versa. Descriptive statistics reinforce the point: all three tools mark most reviews as positive (80.94% for VADER, 80.21% for TextBlob, and a striking 91.10% for Flair). Such skew may reflect the generally appreciative tone of professional wine prose, yet it also suggests that generic analyzers overestimate positivity because they miss subtle critiques embedded in specialised language.

To test whether the U.S.-heavy composition of the corpus biases these results, we compared reviews about U.S. wines with those about all other countries combined. The mean compound polarity differs by only 0.037 points (0.478 vs. 0.515), and the correlation with critic score shifts by just 0.06 (0.29 vs. 0.35). These small gaps imply that, within English-language notes, the generic tools do not systematically inflate or deflate sentiment for non-U.S. wines.

Even so, generic analyzers often invert the polarity of prized wine descriptors. Table A1 from the appendix lists ten tasting note fragments that include approving cues a sommelier would value, phrases such as 'light, crisp, and refreshing', 'zesty nose', 'ripe acidity' and 'familiar and likable', but each fragment receives a neutral (0.00) or negative compound score. This mismatch hints at how general lexicons misread oenological language and underscores the need for a domain-tuned sentiment model.

Together, these findings highlight the limitations of off-the-shelf tools for wine discourse and motivate training a domain-specific analyzer.

3.4. Neural Network-Based Data Analyses

We used complex neural networks to classify wine quality based on written descriptions starting from the specialised language extracted from reviews. If a wine's point score is 90 or higher, we classify it as high quality; otherwise, it is labeled as not high quality.

We split the data into three sets: training, validation, and test. First, we set aside 20% of the data for the test set. Then, from the remaining 80%, we took 12.5% for validation (which amounts to 10% of the total data). This setup allows us to train the model on the training set, use the validation set to tune hyperparameters or early stopping, and keep the test set completely held out for the final evaluation and comparison between models.

3.4.1. Hierarchical Attention Network (HAN)

The Hierarchical Attention Network [24] was adapted to better understand the hierarchical structure of documents, capturing both word- and sentence-level representations. Initially, each word in a sentence was converted into an embedding vector using a pre-trained embedding matrix. A bidirectional Gated Recurrent Unit (GRU) processed these embeddings to produce hidden states for each word. Subsequently, word-level attention processes were employed on these hidden states to derive a weighted representation of the sentence, highlighting terms that significantly enhance its meaning.

For each word w_{it} in sentence S_i, we obtained the embedding x_{it} and computed the hidden state h_{it} using a bidirectional GRU. The word context vector u_{it} was calculated as $u_{it} = \tanh(W_w h_{it} + b_w)$ and the attention weights α_{it} were obtained using a softmax function on the dot product with a word-level context vector u_w. The vector of the sentence s_i is calculated as the weighted sum of the hidden states of the word.

The sentence vectors s_i were processed through another bidirectional GRU to capture inter-sentence dependencies and sentence-level attention was applied similarly to obtain the document representation v. Finally, this document vector v was passed through a softmax classifier to produce the probability distribution over classes.

The model uses the Adam optimiser and binary cross-entropy loss.

3.4.2. Spectral-Normalised Hierarchical Attention Network (S-HAN)

The Spectral-Normalised HAN extended the HAN model by incorporating spectral normalisation to improve training stability and generalisation [84]. Spectral normalisation constrains the Lipschitz constant of the network by normalising the weight matrices. Specifically, for each weight matrix W in the network, we computed the largest singular value $\sigma(W)$ using the power iteration method and scaled the weights as $W_{\text{SN}} = \frac{W}{\sigma(W)}$. We apply this normalisation to all weight matrices in the attention mechanisms and GRU layers while maintaining the same overall architecture as HAN but with normalised weights ensuring controlled spectral norms. By implementing spectral normalisation, we prevent the problem of exploding or vanishing gradients and promote smoother optimisation landscapes. The complete implementation is detailed in Algorithm 1.

Algorithm 1 Hierarchical Attention Network with Spectral Normalisation

Require: Document consisting of sentences $\{S_1, S_2, \ldots, S_N\}$
Ensure: Classification probability y
/* Spectral Normalisation Function */
function SPECTRALNORMALIZE(W)
  $u \leftarrow$ Random vector
  for $k \leftarrow 1$ to num_iterations **do**
    $v \leftarrow \frac{W^\top u}{\|W^\top u\|_2}$
    $u \leftarrow \frac{Wv}{\|Wv\|_2}$
  end for
  $\sigma \leftarrow u^\top W v$
  return $\frac{W}{\sigma}$
end function
/* Apply Spectral Normalisation to Attention and Classifier Weights */
Spectrally normalize attention and classifier weights:
$\mathbf{W}_w^{\text{SN}} \leftarrow \text{SpectralNormalize}(\mathbf{W}_w)$
$\mathbf{u}_w^{\text{SN}} \leftarrow \text{SpectralNormalize}(\mathbf{u}_w)$
$\mathbf{W}_s^{\text{SN}} \leftarrow \text{SpectralNormalize}(\mathbf{W}_s)$
$\mathbf{u}_s^{\text{SN}} \leftarrow \text{SpectralNormalize}(\mathbf{u}_s)$
$\mathbf{W}_c^{\text{SN}} \leftarrow \text{SpectralNormalize}(\mathbf{W}_c)$
/* Word Encoder and Attention */
for each sentence S_i **do**
  for each word w_{it} in S_i **do**
    $\mathbf{x}_{it} \leftarrow \text{Embedding}(w_{it})$
  end for
  $(\mathbf{h}_{i1}, \ldots, \mathbf{h}_{iT_i}) \leftarrow \text{BiGRU}(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT_i})$
  for $t \leftarrow 1$ to T_i **do**
    $\mathbf{u}_{it} \leftarrow \tanh(\mathbf{W}_w^{\text{SN}} \mathbf{h}_{it} + \mathbf{b}_w)$
  end for
  $\alpha_{it} \leftarrow \text{Softmax}\left(\mathbf{u}_{it}^\top \mathbf{u}_w^{\text{SN}}\right)$
  $\mathbf{s}_i \leftarrow \sum_{t=1}^{T_i} \alpha_{it} \mathbf{h}_{it}$
end for
/* Sentence Encoder and Attention */
$(\mathbf{h}_1, \ldots, \mathbf{h}_N) \leftarrow \text{BiGRU}(\mathbf{s}_1, \ldots, \mathbf{s}_N)$
for $i \leftarrow 1$ to N **do**
  $\mathbf{u}_i \leftarrow \tanh(\mathbf{W}_s^{\text{SN}} \mathbf{h}_i + \mathbf{b}_s)$
end for
$\beta_i \leftarrow \text{Softmax}\left(\mathbf{u}_i^\top \mathbf{u}_s^{\text{SN}}\right)$
$\mathbf{v} \leftarrow \sum_{i=1}^{N} \beta_i \mathbf{h}_i$
/* Classification */
$y \leftarrow \text{Softmax}\left(\mathbf{W}_c^{\text{SN}} \mathbf{v} + \mathbf{b}_c\right)$
return y

Spectral normalisation helps neural networks generalise better in several ways. It does this by limiting how much each layer can change its output based on input changes. This makes the model more stable and less likely to overreact to small differences in data. It also acts like a brake on the network's weights, keeping them from getting too big. This helps prevent overfitting, where a model learns the training data too well and cannot handle new information.

3.4.3. Long Short-Term Memory Network (LSTM) with GloVe Embeddings

The LSTM model [23] can capture long-term dependencies within the text by sequentially processing the input sequence of words.

As shown in Algorithm 2, our implementation uses pre-trained GloVe embeddings [85] to enhance the model's understanding of word meanings and similarities. The wine reviews dataset, which includes labels showing the rating of each wine and textual descriptions of the wines themselves, is still the main data utilised to train an LSTM model. During training, the model takes in this data and learns patterns, attitudes, and language usage relevant to its domain. As a result of rigorous training on massive, general-purpose text corpora, GloVe embeddings give a rich, general-purpose semantic representation of words. These corpora include Common Crawl and Wikipedia. With the help of the GloVe embeddings, which provide the model with pre-existing knowledge about word meanings and associations, the model can optimise its parameters to minimise loss on the wine reviews dataset. By learning patterns particular to wine reviews, training on the data, and fine-tuning the embeddings, the model becomes domain-specific. Starting with pre-trained, embeddings provides a solid foundation for the model, often leading to faster convergence during training and potentially better performance on general language aspects. Furthermore, by keeping the embedding layer trainable and handling Out-of-Vocabulary (OOV) words appropriately, the model can adapt the pre-trained embeddings to better fit the wine reviews' domain.

An embedding matrix is created for the model's vocabulary, initialising the embeddings with corresponding GloVe vectors. The model then maps each word in the input sequence to its embedding vector, creating a sequence of embedding vectors that represent the semantic content of the text. The sequence is fed into a bidirectional LSTM network, which updates its hidden and cell states based on the current input and previous states. The model computes the mean of all hidden states across the sequence, providing a comprehensive representation that captures the overall context. Next, it is trained using a suitable loss function and an optimiser (i.e., Adam), ensuring it does not become biased toward the majority class. The Adam optimisation algorithm [86] is an extension to stochastic gradient descent that has gained popularity for deep learning applications, including natural language processing. It is designed to update network weights iteratively based on training data, offering several benefits such as simplicity, computational efficiency, minimal memory requirements, and adaptability to diagonal rescaling of gradients. This approach presented offers several advantages, including semantic enrichment, improved generalisation, fine-tuning, domain-specific vocabulary, mean pooling over hidden states, and robustness to vocabulary gaps.

Algorithm 2 Long Short-Term Memory Network with GloVe Embeddings

Require: Input sequence of words $X = (w_1, w_2, \ldots, w_T)$
Ensure: Classification probabilities $\hat{y}$
/* Initialize Embedding Layer with GloVe Embeddings */
Load pre-trained GloVe embeddings $\mathbf{E}_{\text{GloVe}}$
Build vocabulary $\mathcal{V}$ from training data
Initialize embedding matrix $\mathbf{E} \in \mathbb{R}^{|\mathcal{V}| \times d}$ where d is embedding dimension
for each word $w \in \mathcal{V}$ **do**
  if w in GloVe vocabulary **then**
    $\mathbf{E}(w) \leftarrow \mathbf{E}_{\text{GloVe}}(w)$ ▷ Assign pre-trained GloVe embedding
  else
    Initialize $\mathbf{E}(w)$ randomly ▷ For words not in GloVe
  end if
end for
/* Embedding Layer */
for $t \leftarrow 1$ to T **do**
  $\mathbf{x}_t \leftarrow \mathbf{E}(w_t)$ ▷ Map word w_t to embedding vector $\mathbf{x}_t$
end for
/* LSTM Encoder */
Initialize hidden state $\mathbf{h}_0$ and cell state $\mathbf{c}_0$ to zero vectors
for $t \leftarrow 1$ to T **do**
  $\mathbf{f}_t \leftarrow \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f)$ ▷ Forget gate
  $\mathbf{i}_t \leftarrow \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i)$ ▷ Input gate
  $\tilde{\mathbf{c}}_t \leftarrow \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c)$ ▷ Candidate cell state
  $\mathbf{c}_t \leftarrow \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t$ ▷ Update cell state
  $\mathbf{o}_t \leftarrow \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o)$ ▷ Output gate
  $\mathbf{h}_t \leftarrow \mathbf{o}_t \odot \tanh(\mathbf{c}_t)$ ▷ Update hidden state
end for
/* Sequence Representation */
Compute mean of hidden states: $\mathbf{h}_{\text{mean}} \leftarrow \frac{1}{T} \sum_{t=1}^{T} \mathbf{h}_t$
/* Classification Layer */
Apply dropout: $\mathbf{h}_{\text{drop}} \leftarrow \text{Dropout}(\mathbf{h}_{\text{mean}})$
Compute logits: $\mathbf{z} \leftarrow \mathbf{W}_{\text{cls}} \mathbf{h}_{\text{drop}} + \mathbf{b}_{\text{cls}}$
Compute class probabilities: $\hat{y} \leftarrow \text{Softmax}(\mathbf{z})$
return $\hat{y}$

3.4.4. Hierarchical Attention Network (HAN) with GloVe Embeddings

The LSTM + GloVe approach previously discussed treats the entire text as a sequence of words and processes it using an LSTM network, capturing temporal dependencies but potentially struggling with long documents due to issues like vanishing gradients and computational complexity. To address these problems, in the algorithm presented in Algorithm 3, we also used GloVe embeddings to improve the HAN algorithm. The semantic richness of GloVe embeddings is useful for both LSTM and HAN models, but the latter's design makes it possible for these embeddings to engage with attention mechanisms on several levels, which can result in more accurate feature representations.

Tuning hyperparameters and allocating computational resources with precision is necessary for HAN because of its increased complexity caused by extra GRU layers and attention processes. Nevertheless, compared to the LSTM method, the model's increased complexity allows it to grasp more subtle data patterns, which frequently leads to improved generalisation and improved classification accuracy.

We can think of GloVe embeddings as a language processing agent that has learned a great deal about the links between words and their meanings via studying a wide variety of texts. Although it lacks subject expertise, this agent has a general grasp of language semantics. The agent can concentrate on and learn domain-specific information with the help of the HAN, which functions as an adaptive mechanism while processing specialised content such as wine evaluations. Words like "tannin," "bouquet," "oak," and "finish" are specific to wine evaluations and are essential for the agent to recognise and

highlight in order to grasp the finer points of the text. To determine the overall emotion or quality evaluation of the wine, the AI can prioritise sentences at the phrase level using the attention mechanism. For example, words that emphasise notable qualities or offer comparative assessments are given greater weight. The agent improves its general language representations to match the specialised vocabulary and stylistic expressions of the wine industry by training on the specific dataset of wine reviews. By providing a hierarchical framework for the texts, the HAN makes it easier for the agent to adapt by directing its attention to the most informative words and sentences.

Algorithm 3 Hierarchical Attention Network with GloVe Embeddings

Require: Document consisting of sentences $\{S_1, S_2, \ldots, S_N\}$
Ensure: Classification probabilities $\hat{y}$
/* Load Pre-trained GloVe Embeddings */
Load GloVe embedding matrix $\mathbf{E}_{\text{GloVe}} \in \mathbb{R}^{|\mathcal{V}| \times d}$
/* Initialize Embedding Layer */
Build vocabulary $\mathcal{V}$ from training data
Initialize embedding matrix $\mathbf{E} \in \mathbb{R}^{|\mathcal{V}| \times d}$
for each word $w \in \mathcal{V}$ **do**
  if w in GloVe **then**
    $\mathbf{E}(w) \leftarrow \mathbf{E}_{\text{GloVe}}(w)$
  else
    Initialize $\mathbf{E}(w)$ randomly from $\mathcal{U}(-0.25, 0.25)$
  end if
end for
/* Word Encoder and Attention */
for each sentence S_i **do**
  for each word w_{it} in S_i **do**
    $\mathbf{x}_{it} \leftarrow \mathbf{E}(w_{it})$ ▷ Embed word
  end for
  $\{\mathbf{h}_{i1}, \ldots, \mathbf{h}_{iT_i}\} \leftarrow \text{BiGRU}(\{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT_i}\})$
  for $t \leftarrow 1$ to T_i **do**
    $\mathbf{u}_{it} \leftarrow \tanh(\mathbf{W}_w \mathbf{h}_{it} + \mathbf{b}_w)$
  end for
  Compute word attention weights:
  $\alpha_{it} \leftarrow \dfrac{\exp(\mathbf{u}_{it}^\top \mathbf{u}_w)}{\sum_{t'=1}^{T_i} \exp(\mathbf{u}_{it'}^\top \mathbf{u}_w)}$
  Compute sentence representation: $\mathbf{s}_i \leftarrow \sum_{t=1}^{T_i} \alpha_{it} \mathbf{h}_{it}$
end for
/* Sentence Encoder and Attention */
$\{\mathbf{h}_1, \ldots, \mathbf{h}_N\} \leftarrow \text{BiGRU}(\{\mathbf{s}_1, \ldots, \mathbf{s}_N\})$
for $i \leftarrow 1$ to N **do**
  $\mathbf{u}_i \leftarrow \tanh(\mathbf{W}_s \mathbf{h}_i + \mathbf{b}_s)$
end for
Compute sentence attention weights:
$\beta_i \leftarrow \dfrac{\exp(\mathbf{u}_i^\top \mathbf{u}_s)}{\sum_{j=1}^{N} \exp(\mathbf{u}_j^\top \mathbf{u}_s)}$
Compute document representation: $\mathbf{v} \leftarrow \sum_{i=1}^{N} \beta_i \mathbf{h}_i$
/* Apply Dropout */
$\mathbf{v}_{\text{drop}} \leftarrow \text{Dropout}(\mathbf{v},\ p = 0.5)$
/* Classification Layer */
Initialize classifier weight $\mathbf{W}_c$ using Xavier initialisation
Compute logits: $\mathbf{z} \leftarrow \mathbf{W}_c \mathbf{v}_{\text{drop}} + \mathbf{b}_c$
Compute class probabilities: $\hat{y} \leftarrow \text{Softmax}(\mathbf{z})$
return $\hat{y}$

3.5. Extracting Wine Characteristics and Generating Marketing Images

Next, we employed a multistage natural language processing approach to extract salient characteristics from high-scoring wine reviews and generate image prompts for

AI-based visualisation to be used for marketing purposes. The approach is shown in Algorithm 4.

Algorithm 4 Wine Review Image Prompt Generation

Require: DataFrame D containing wine reviews and scores
Ensure: Set of image prompts P
Parameters:
N: Number of top wines to consider
K: Number of top characteristics to extract
M: Number of prompts to generate
/* Extract Top Wine Reviews */
1: $R \leftarrow \text{TopN}(D, \text{key='points'}, N)$
/* Extract Characteristics */
2: $C \leftarrow \emptyset$
3: **for** each review $r \in R$ **do**
4: $W \leftarrow \text{Tokenize}(r)$
5: $B \leftarrow \text{ExtractBigrams}(W)$
6: $T \leftarrow \text{POSTag}(W)$
7: $C \leftarrow C \cup \{w \mid (w, \text{tag}) \in T, \text{tag} \in \{\text{'JJ'}, \text{'NN'}\}\} \cup B$
8: **end for**
/* Select Top Characteristics */
9: $C_{\text{top}} \leftarrow \text{TopK}(\text{FrequencyCount}(C), K)$
/* Generate Prompts */
10: $P \leftarrow \emptyset$
11: **for** $i \leftarrow 1$ to M **do**
12: $C_{\text{sample}} \leftarrow \text{RandomSample}(C_{\text{top}}, \min(8, |C_{\text{top}}|))$
13: $p \leftarrow \text{GeneratePrompt}(C_{\text{sample}})$
14: $P \leftarrow P \cup \{p\}$
15: **end for**
16: **return** P
17: **function** GENERATEPROMPT(C_{sample})
18: base ← "Create a wine image with the following characteristics:"
19: elements ← $\emptyset$
20: **for** each characteristic $c \in C_{\text{sample}}$ **do**
21: elements ← elements ∪ {"Visual element or quality evoking '{c}'"}
22: **end for**
23: elements ← elements ∪ {
"Elegant lighting accentuating color and texture.",
"Background suggesting luxury and sophistication.",
"High-quality, photorealistic style."
24: }
25: **return** base + Join(elements)
26: **end function**

As shown in Algorithm 4, the top N wines, ranked by their point scores, were selected for analysis to ensure a focus on the characteristics of premium wines (line 2). Each review was tokenised (line 6), followed by bigram extraction (line 7), and part-of-speech tagging (line 8). The algorithm specifically selects adjectives (tagged as JJ) and nouns (tagged as NN), which usually carry the most descriptive information in wine reviews, along with the extracted bigrams (line 9).

The extracted characteristics, including individual words and bigrams, were combined from all reviews (lines 4–10). These characteristics were then ranked by frequency and the K most common characteristics were selected to represent the core attributes of premium wines in the dataset (line 12).

To generate prompts, we randomly sampled eight characteristics from the top K (line 15). A semi-structured template for image generation prompts was developed in the GeneratePrompt function (lines 17–26), incorporating the sampled characteristics and additional descriptive elements.

The algorithm does not explicitly categorise characteristics into predefined oenological dimensions (colour, body, taste, and aroma). Instead, it treats all characteristics equally, focusing on their frequency and random sampling for diversity (C_{sample}). Then, it generates M unique prompts (lines 13–17), each designed to guide an AI image generation tool in creating visually compelling representations of premium wines for marketing purposes.

The GeneratePrompt function creates a base prompt and then adds visual elements for each characteristic sampled (lines 20–22). It also includes standard elements about lighting, background, and image quality (lines 23–24).

By sampling from the top K characteristics, the approach ensures variability in the generated prompts while maintaining a focus on the most relevant characteristics of premium wines, as determined by their frequency in top-rated wine reviews.

3.6. Evaluating the Generated Marketing Images

Inspired by the approach of Oppenlaender et al. (2024) [87], we first sampled 20 diverse, wine-specific prompts. Each prompt was rendered once by ChatGPT 4o Image Generation and, using the identical text, once by Midjourney v6. ChatGPT returns a single image per prompt, whereas Midjourney outputs four variants; we consistently selected the first variant to avoid cherry-picking, yielding a corpus of 40 images (20 per generator).

For every image, participants later provided two 5-point Likert ratings: relevance (brief-fit) and appeal (commercial viability). After completing all image ratings, participants indicated their perceived stylistic difference between the two sets, a general preference, the factors that influenced their judgments (visual appeal, creativity, emotional appeal, brand/commercial viability, technical quality, realistic presentation), and basic demographics, including self-reported wine-marketing experience (1 = none, 5 = expert).

The evaluation ran on a custom Python/Flask website hosted on an Ubuntu virtual machine (Oracle Cloud) with Gunicorn, Nginx and Let's Encrypt SSL. Gunicorn served the Flask app concurrently to multiple users, Nginx handled static content, and HTTPS encryption ensured secure transmission of responses.

The Web page consisted of four sequential pages: (1) participant information, (2–3) evaluation of Image Sets 1 and 2, and (4) demographic and post-evaluation questions. For each image, participants provided ratings on two 5-point Likert scales: relevance (how well the image matches the brief) and appeal (commercial viability for advertising). The images were presented with their corresponding briefs and the AI source was not disclosed to ensure blind evaluation.

Each session received a unique ID; all responses were stored row-wise in a comma-separated file with 91 columns (demographics, post-task items, and per-image ratings). The Web page's source code and the 40 generated images are openly available in the GitHub repository cited in the Data Availability Statement.

4. Results

For RQ1, we correlated the three-level style difference scale (*1 ="Very similar"*, *3 ="Very different"*) with preference strength and with set-level rating gaps using Pearson's r. For RQ2, we decomposed binary factor checkboxes and ran point-biserial correlations plus logistic regressions featuring the interaction term (factor $\times$ expertise). For RQ3, we benchmarked four classifiers—HAN, Spectral-HAN, HAN + GloVe, and LSTM + GloVe. Accuracy, precision, recall, and macro-F_1 are summarised in Table 6, while paired McNemar tests on held-out errors appear in Table 7. Validation-epoch curves are plotted in Figures 2 and 3. For RQ4, we compared a generic sentiment tool with our wine-tuned lexicon on a manually labelled subset of 1200 sentences. Polarity accuracy and macro-F_1 are reported in Table 5; tool characteristics are contrasted in Table 4.

All analyses were executed in Python 3.9.12 (pandas 2.2.3, scipy 1.13.1, statsmodels 0.15.0).

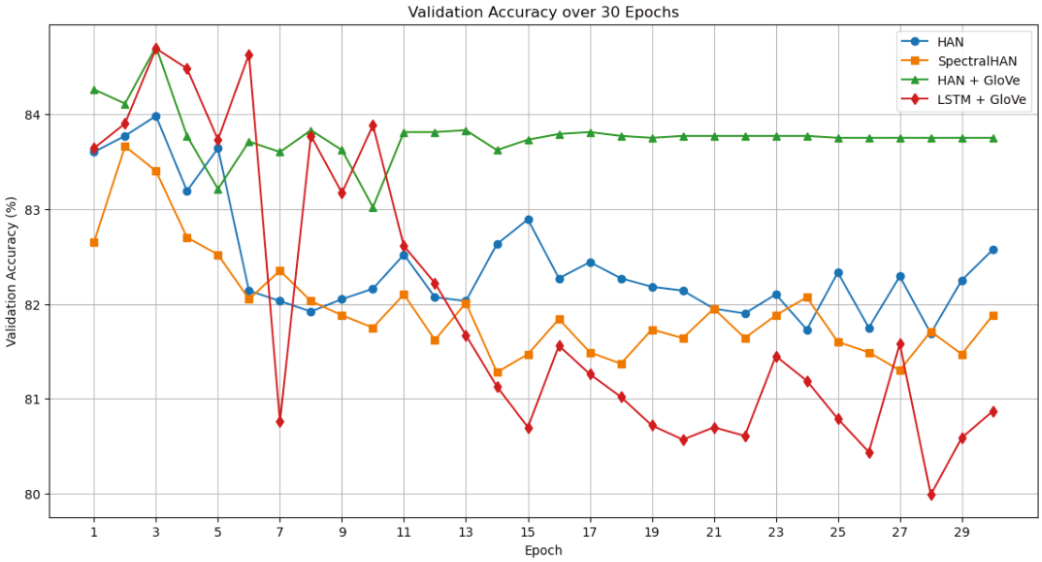


Figure 2. Validation accuracy over epochs.

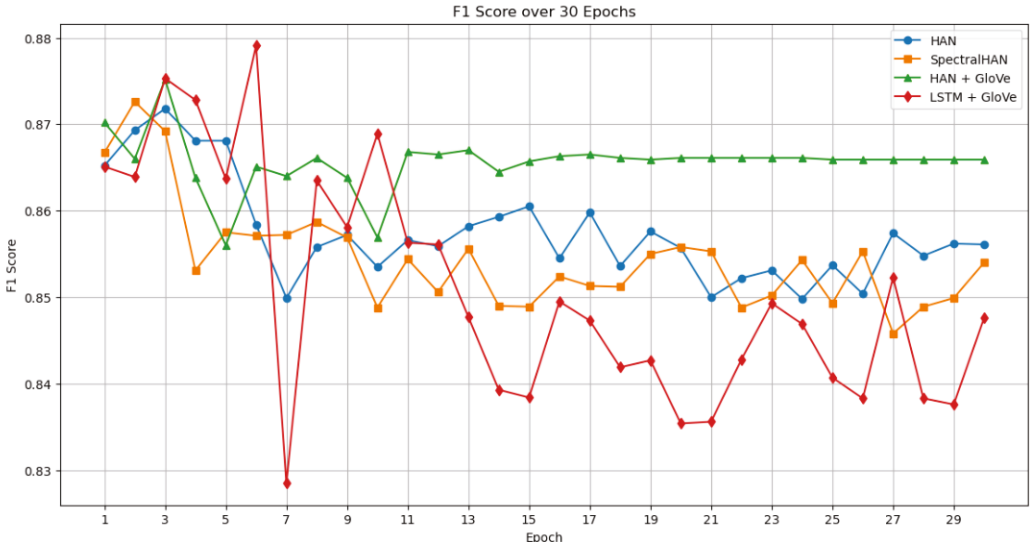


Figure 3. F1 score across epochs.

Table 6. Performance comparison of different models.

Metric	HAN	SpectralHAN	HAN + GloVe	LSTM + GloVe
Init. Acc. (Ep.1)	83.60%	82.65%	**84.26%**	83.64%
Best Acc.	83.98% (Ep.3)	83.66% (Ep.2)	**84.71%** (Ep.3)	84.69% (Ep.3)
Final Acc. (Ep.30)	**82.57%**	81.88%	83.75%	80.87%
Best Precision	86.67% (Ep.1)	88.40% (Ep.4)	89.66% (Ep.5)	**90.73%** (Ep.7)
Best Recall	89.33% (Ep.3)	**92.53%** (Ep.1)	87.95% (Ep.3)	91.69% (Ep.6)
Best F1	87.18% (Ep.3)	87.26% (Ep.2)	**87.52%** (Ep.3)	**87.91%** (Ep.6)
Final Precision	**86.13%**	83.92%	87.13%	82.38%
Final Recall	85.10%	**86.93%**	86.05%	87.28%
Final F1	**85.61%**	85.40%	86.59%	84.76%

Note: Bold values show the best values for the metric.

Table 7. Summary of McNemar's test results comparing model performance.

Model Comparison	Disagree	Test Stat	p-Value	Acc. Diff	Signif.	Conclusion
HAN vs. S-HAN	14.41%	0.007	0.935	0.04%	No	No significant difference
HAN vs. LSTM	16.06%	14.054	**0.0002**	1.57%	Yes	HAN better than LSTM
HAN vs. HAN + GloVe	13.20%	10.017	**0.002**	1.20%	Yes	HAN + GloVe better than HAN
S-HAN vs. LSTM	15.20%	15.679	**0.0001**	1.61%	Yes	S-HAN better than LSTM
S-HAN vs. HAN + GloVe	12.88%	9.541	**0.002**	1.16%	Yes	HAN + GloVe better than S-HAN
LSTM vs. HAN + GloVe	12.43%	57.037	**<0.0001**	2.77%	Yes	HAN + GloVe better than LSTM

Note: Significance determined by $p < 0.05$. S-HAN = Spectral-Normalised HAN. "Disagree" = percentage of samples with different predictions. Bold values for $p < 0.05$.

4.1. HAN Analyses

We evaluated the performance of four different neural network models on a wine classification task: the Hierarchical Attention Network (HAN), the Spectral Hierarchical Attention Network (SpectralHAN), HAN augmented with pre-trained GloVe embeddings (HAN + GloVe), and a Long-Short-Term Memory network with GloVe embeddings (LSTM + GloVe). The dataset was divided into a training set of 32,602 samples, a validation set of 4658 samples, and a test set of 9316 samples, with a vocabulary size of 6461 unique tokens.

Table 6 summarises the key performance metrics for each model, including accuracy, precision, recall, and F1 score across 30 training epochs. Figures 2 and 3 illustrate the trends in validation accuracy and F1 scores over the training epochs for all models.

The learning patterns and performance characteristics of four models were evaluated on a wine classification task: Hierarchical Attention Network (HAN), SpectralHAN, HAN enhanced with GloVe embeddings (HAN + GloVe), and LSTM with GloVe embeddings (LSTM + GloVe). With a dataset of 4658 validation samples, 9316 test samples, and 32,602 training samples, we tracked how each model evolved and became more generalised over training.

HAN showed consistent performance, peaking at 83.98% validation accuracy early in training and staying relatively close to this level all the way to 82.57% at the end. SpectralHAN had a similar pattern, peaking early at 83.66% accuracy and finishing at 81.88%. HAN + GloVe achieved 84.71% peak accuracy, the highest of all models, and continued to perform well, completing at 83.75%. On the other hand, LSTM + GloVe equaled HAN + GloVe's early performance, hitting 84.69%, but then showed a significant downturn in the last stages, finishing at 80.87%.

HAN + GloVe consistently provided high precision, peaking at 89.66% and remaining around 87% near the conclusion, indicating trustworthy positive identifications. LSTM + GloVe reached the highest precision of 90.73%; however this dropped to 82.38% in later epochs. SpectralHAN originally had the highest recall of 92.53%, indicating a good capacity to recognise positive cases early in training. The F1 scores for all models peaked between the second and sixth epochs, with HAN + GloVe and LSTM + GloVe both above 87%, indicating a great balance of precision and recall throughout the early training periods.

Notably, HAN + GloVe demonstrated exceptional stability throughout the training phase. There was minimal change in its validation accuracy and F1 score in later epochs after initial improvements peaked. The metrics of the HAN + GloVe model don not change after epoch 17, as we can see. Since the StepLR scheduler of the training code is set to decrease the learning rate by 0.1 every five epochs from an initial learning rate of 0.001, the HAN + GloVe model performance metrics hit a plateau after epoch 17. This indicates that the learning rate undergoes an exponential decay as the training progresses. It stays at 0.001 for the first five epochs, drops to 0.0001 for the sixth to tenth, rises to 0.00001 for the eleventh to fifteenth, and then continues to fall by an order of magnitude every five

epochs. The learning rate drops to 0.000001 by Epoch 16, and it reaches 0.00000001 at epoch 26. Because the optimiser has very low learning rates, after Epoch 15 it is essentially updating the model's parameters insignificantly. Since the gradient descent steps are so small, they have little effect on the model's weights, so the model's accuracy and loss are essentially constant throughout subsequent epochs. The training process comes to a standstill because the optimiser cannot improve the model's performance by adjusting the parameters correctly. Despite our best efforts, no matter how we tweaked the decay factor, increased the step size, or switched to a different scheduler (ReduceLROnPlateau), the top values of the model metrics did not improve; instead they deteriorated.

Overall, the combination of hierarchical attention and domain-informed embeddings proved to be the most balanced and dependable option. This synergy allowed the model to focus on the most telling words and phrases in each review while also taking advantage of broader contextual clues captured by GloVe vectors. Put simply, the attention mechanism's ability to highlight especially meaningful segments of text worked hand-in-hand with the embeddings' capacity to encode semantic relationships learned from the corpora.

In the context of the analysis, McNemar's test was employed on the hold-out (validation) set to determine whether there is a statistically significant difference in the classification accuracies of the models. By comparing the models' predictions on the same set of wine quality instances, the test evaluates whether one model consistently outperforms the other. The results are presented in Table 7.

The results of Table 7 show that HAN + GloVe consistently performs better than the other models, proving that the improved accuracy of the model is statistically significant and practically relevant for our task. The incorporation of GloVe embeddings improves the model's precision in categorising wine reviews. Even though all models demonstrated strengths in capturing nuances of wine language, it was the hierarchical attention approach, when supported by rich pretrained embeddings, that consistently rose to the top.

4.2. Comparison of the Classification Results

Table 8 juxtaposes our Hierarchical Attention Network with GloVe embeddings (HAN + GloVe) against the CNN, BiLSTM, and BERT pipelines of Katumullage et al. [22] (their models were trained on ∼140 k *Wine Spectator* notes, whereas we use 46 k *Wine Enthusiast* reviews; absolute accuracies are therefore not directly comparable) and a multilingual DistilBERT baseline fine-tuned on our corpus.

Table 8. Performance comparison on the *Wine Enthusiast* test split ($n = 9316$). The macro-averaged precision, recall, and F1 are reported.

Model	Accuracy	Precision	Recall	F1 Score
HAN + GloVe (ours)	84.71	**89.66**	86.97	**88.29**
CNN [22]	88.02	87.08	86.26	86.67
BiLSTM [22]	88.69	87.59	87.30	87.44
BERT [22]	**89.12**	87.97	**87.79**	87.88
mDistilBERT (baseline)	84.97	84.93	84.97	84.95

Note: Bold values show the best values for the metric.

Although its accuracy trails the heavier CNN/BiLSTM/BERT pipelines, our model achieves the highest precision (89.7%) and overall F_1 (88.3%), edging out the BERT baseline by 0.4 pp. In the pipeline from Figure 1 the classifier acts as a gatekeeper: with a 0.5 decision threshold it discards 41% of low-scoring reviews while retaining 92% of those rated ≥90 points, thus supplying a high-purity pool for keyword extraction and prompt generation. Our HAN + GloVe model contains ≈6 million parameters (about 4% of the 135 million in multilingual DistilBERT), achieves a comparable F_1 (88.3% vs. 84.9%),

and processes a review ≈4× faster on a T4 GPU (9 ms vs. 38 ms) while still providing interpretable attention weights. These weights appear as two vectors—word–level α_{it} and sentence-level β_i—which can be visualised as a heat-map overlay on the review text; marketers instantly see which words and sentences drove the prediction and can lift the top-weighted descriptors directly into their creative briefs. In high-stakes curation tasks, explainability can outweigh the small numeric margin.

4.3. Wine Characteristic Extraction

The algorithm presented in Section 3.5 extracted and analyzed the top characteristics from the wine reviews, yielding the following 15 most frequent terms: wine, tannins, acidity, palate, black, time, drink, years, magnificent, fruits, structure, leather, red, aromas, and new. These terms represent a mix of sensory descriptors, temporal references, and structural elements commonly associated with high-quality wines.

Using these characteristics, the system generated multiple prompts. For example,

> "Create a captivating wine advertisement image featuring a glass of deep red, full-bodied wine. The wine should appear rich and fruity. Include elegant lighting that accentuates the wine's colour and texture. In the background, subtly showcase elements that suggest luxury and sophistication. Incorporate visual elements or symbols that evoke the sense of acidity. Incorporate visual elements or symbols that evoke the sense of fruits. Incorporate visual elements or symbols that evoke the sense of drink. The overall style should be high-quality, photorealistic, and visually striking."

All prompts consistently included instructions for elegant lighting, luxury elements in the background, and a high-quality photorealistic style.

Figures 4 and 5 present sample images generated using the same prompts on ChatGPT and Midjourney, respectively, allowing a direct comparison between the two AI image generation systems.

Figure 4. Selection of the generated images—ChatGPT.

Figure 5. Selection of the generated images—Midjourney.

These visual examples show the distinct artistic styles and interpretation approaches employed by each platform when processing the same textual descriptions. All forty images can be found in the wine-evaluation (https://github.com/vladdiac/wine-evaluation, branch main, commit 1f3a964, 13 June 2025) and winereviews (https://github.com/vladdiac/winereviews, branch main, commit 27e9d97, 13 June 2025) GitHub repositories mentioned in the Data Availability Statement at the end of the article.

4.4. Image Evaluation

Participants were asked whether the two image sets looked "very similar," "somewhat different," or "very different." Table 9 reports, for each category.

- *n*—number of participants;
- % Set 2 pref.—percentage whose preference strength was *above zero* (they chose "prefer Set 2" or "strongly prefer Set 2");
- Preference strength—mean on a five-point scale (−2 = strongly Set 1 . . . +2 = strongly Set 2);
- Δ relevance and Δ appeal—the average rating difference (Set 2 minus Set 1); positive values favour Set 2, negative values favour Set 1.

Table 9. Preference and rating gaps by perceived stylistic difference.

Style Difference	*n*	% Set 2 Pref.	Pref. Strength	Δ Relevance	Δ Appeal
Somewhat different	16	31.25	−0.25	−0.15	−0.23
Very different	8	87.50	0.88	−0.09	0.19
Very similar	5	20.00	−0.80	0.18	−0.03

Table 9 makes the pattern explicit. When participants judged the two image sets as *very different* ($n = 8$), 88 % expressed at least a mild preference for Set 2 and their average preference strength was +0.88 on the −2 to +2 scale. By contrast, those who found the

sets *very similar* ($n = 5$) preferred Set 2 only 20 % of the time and scored -0.80 on the same scale. The resulting Pearson correlation of $r = 0.49$ ($p = 0.007$) indicates a moderate positive association. As participants moved one step up the three-level style difference scale (for example, from 'very similar' to 'somewhat different'), their mean preference score increased by about 0.5 points on the five-point scale.

For the expertise analysis, the self-reported wine-marketing scale (1 = none, 5 = expert) was split into novices (1–2) and experts (3–5). Table 10 shows, for each rating factor,

- Endorsement rates (percentage ticking the factor) in novices and experts;
- Point-biserial correlation r between endorsing the factor (1 = yes, 0 = no) and higher appeal scores within each group (positive r = higher appeal when the factor is cited).

Table 10. Factor endorsement rates and correlations with appeal.

Factor	% Novices	% Experts	r (Novices)	r (Experts)
Visual appeal/aesthetics	83.3	90.9	0.20	−0.29
Creativity/originality	50.0	36.4	0.25	0.16
Emotional appeal	38.9	27.3	0.05	0.49
Brand/commercial viability	16.7	27.3	−0.23	−0.17
Technical image quality	44.4	45.5	−0.30	−0.10
Realistic product presentation	55.6	45.5	−0.13	−0.43

Using the five-point experience item, participants were split into novices (levels 1–2, $n = 13$) and experts (levels 3–5, $n = 16$). Table 10 shows that novices mentioned creativityin 50 % of cases and, when they did, their mean appeal ratings rose (point biserial $r = 0.25$). Experts endorsed creativity less often (36%) and showed almost no association with appeal ($r = 0.16$). Instead, experts in our sample penalised technical lapses ($r = -0.30$) and brand-fit issues ($r = -0.17$), whereas novices were largely insensitive to those factors. The factor-by-expertise interaction therefore confirms that the criteria driving appeal shift markedly with professional experience.

5. Discussion

In this study, we used a Hierarchical Attention Network (HAN) with GloVe embeddings to handle the nuanced language of wine reviews. Although HANs and pre-trained embeddings have been employed in other text classification tasks, this work underscores their utility in capturing fine-grained wine descriptors for improved classification. The model processes text in a hierarchical way, analyzing both words and sentences, to pinpoint the terms most indicative of quality, style, or flavour. Drawing on GloVe word embeddings allows the classifier to align closely with domain-specific nuances such as "tannin," "oak," or "citrus," improving its ability to map textual clues to meaningful wine attributes. This hierarchical attention captures context more effectively than traditional flat models. For example, word- and sentence-level attention weights focus on the multiword cue 'ripe black cherry at the end', while a bag-of-words baseline would allocate equal weight to each individual token. By incorporating pre-trained GloVe embeddings, the model also benefits from rich semantic knowledge; wine-related terms are understood in context (e.g., "dry" vs. "sweet" or "Bordeaux" vs. "Rioja"), even if those exact word combinations never appeared in the training data. This combination of HAN + GloVe thus provides both accuracy and interpretability. It outperforms simpler classifiers (like bag-of-words or basic LSTMs) by a substantial margin, aligning with what prior studies have found for HANs on general text tasks. Equally important, the attention weights offer guidance that a sommelier or marketer can appreciate—the model can explain its

predictions by highlighting tasting notes that drove the classification, which adds trust and transparency to the AI's decisions.

In our data, reviews from the United States make up the largest single country segment (46.5 percent) and almost all New World reviews (see Table 3), but overall there is a nearly even split between Old World (European) and New World wine regions, at around 47 percent versus 53 percent. However, even though the geographic distribution is almost balanced, there are some stylistic differences that can affect the performance of the model. In particular, Old World regions often produce more sparkling wines, which come with their own unique vocabulary (terms such as "mousse," "autolytic," "perlage," and "dosage"). Meanwhile, New World reviews can be heavily skewed toward still wines, especially bold reds. Sparkling wine descriptors appear in only seven to eight percent of reviews, although sparkling wines account for a significant portion of global wine production, especially in the Old World.

Our error analysis showed a slight accuracy gap of just under 2% in favour of New World wines, suggesting that the model largely captures the distinct language patterns of both traditions. Still, the under-representation of sparkling wine terminology may limit the model's ability to accurately classify these wines or detect key quality markers.

If necessary, we could address this by using style-based sampling in addition to regional considerations so we capture a wider range of wine types. This approach would help ensure that specialised categories such as sparkling wines, which have their own unique set of descriptors, receive sufficient attention in the model.

5.1. Research Questions: Synthesis and Key Insights

RQ1 hints that visual distinctiveness resonates with consumers. As Table 9 shows, larger style gaps boost both absolute preference and the rating spread between image sets. This echoes the multi-domain sentiment gain that hierarchical context delivers in textual classification [88] and aligns with Cao et al.'s U-shaped congruence curve, where either perfect fit or deliberate dissonance drives choice [89]. Together, these results imply that clear visual–verbal signalling—whether via distinctive style or hierarchical language cues—helps consumers differentiate products.

RQ2 suggests that aesthetics lure novices; authenticity reassures experts. Point-biserial tests in Table 10 rank "label design" as the top novice driver, whereas the significant factor $\times$ expertise interaction shows experts pivot to authenticity cues. Wine, perfume, and food reviews indeed use largely non-overlapping sensory vocabularies, each tuned to its domain [90], and a graph-neural model that injects external lexica lifts aspect accuracy in a similar way [91]. Luo et al. further show that a visual sweetness scale helps novices choose but leaves experts unmoved [92], underscoring how cue utilisation shifts with knowledge.

Hierarchical context demonstrates clear advantages over depth alone, as evidenced in our analysis of RQ3. While LSTM+GloVe achieves superior peak precision (90.73%) and F1 scores (87.91%) in Table 6, HAN+GloVe consistently delivers the highest classification accuracy (84.71%) and demonstrates significantly better error patterns in paired McNemar tests (Table 7), outperforming LSTM + GloVe by 2.77% with statistical significance ($p < 0.0001$). The learning curves (Figures 2 and 3) reveal that HAN + GloVe maintains more stable performance throughout training epochs, suggesting better generalisation despite not achieving the absolute peak precision metrics. This stability, combined with the interpretable attention weights that highlight domain-specific wine descriptors, makes HAN+GloVe preferable for practical deployment where consistent accuracy and explainability matter more than peak precision on individual metrics. A wine-specific BERT study likewise predicts style and quality directly from expert notes, reinforcing the value of structured context [93].

By addressing RQ4 it is clear that generic tools capture only a slice of wine sentiment. Table 5 shows that VADER achieves the highest correlation with expert points ($r = 0.32$) yet still explains barely 10% of the variance, while TextBlob's correlation is almost negligible ($r = 0.13$). All three tools are heavily skewed toward positive polarity (81–91% positive), suggesting they mistake descriptive sensory terms for praise. Such mis-scoring of words like dry or tannin echoes the lexicon-mismatch problem flagged by Hörberg et al. in cross-domain vocabularies [90]. These results reinforce the need for a domain-tuned lexicon or an external-knowledge model of the kind Gu et al. deploy for aspect sentiment [91], which we leave as a next-step upgrade to our pipeline.

Generative AI can outperform human creatives when its output looks natural [7], yet acceptance drops if images feel "eerie" unless balanced by clear informational value [94]. By chaining hierarchical NLP, domain-tuned sentiment, and quality-checked AI visuals, our pipeline offers a roadmap that satisfies both novices and connoisseurs—advancing the generative-marketing frontier sketched by Cillo and Rubera [95].

5.2. E-Commerce Applications and Integration into AI Driven Architecture

From an industry perspective, our approach has significant implications for online wine retail and e-commerce. An immediate benefit is automated product classification and tagging. Wine retailers often deal with thousands of labels and vintages; manually curating descriptions and categories is labor intensive. Our HAN-based classifier can automatically assign wines to flavour profiles, regions, or style categories based on their descriptions, ensuring consistent metadata across the catalog. This kind of auto-tagging streamlines inventory management. Therefore, as soon as a new wine is added to the database with its description, the model can classify it (e.g., "full-bodied red", "Tuscany", "notes of vanilla and cherry") within seconds, with accuracy that approaches expert human level. By labeling and organising product data at scale, the system reduces human error and saves countless hours. Beyond operational efficiency, the model greatly enhances personalised shopping experiences. In e-commerce, personalisation is key to customer engagement, and our classifier becomes a critical enabler for a recommendation engine. The rich profile each wine receives (from the classification output) can be matched with user preferences. For example, if a customer consistently buys wines that the model classifies as "bold" and "fruity", the recommendation system can suggest other bottles with similar HAN-derived profiles, even if the user has never heard of them. These AI-driven recommendations go beyond simple collaborative filtering; they are grounded in an understanding of the product content (the wine's taste and style) and the user's taste in language terms. This content-based personalisation means the site can highlight "Because you liked X (which has notes of blackberry and chocolate), we recommend Y, which shares a similar flavour profile." Shoppers thus discover new wines through flavour notes they know they enjoy, increasing satisfaction and trust in the platform. Such tailored recommendations and search results have been shown to increase customer engagement, conversion rates, and loyalty for online retailers. In essence, our model functions as a virtual sommelier on the e-commerce site, automating expertise to guide buyers in a personalised manner.

To obtain these benefits in a real e-commerce setting, we propose an architecture that integrates the HAN + GloVe classifier as part of a larger AI-driven system. This would involve several key components working together.

- Data Ingestion and Preprocessing: First, the system gathers textual data from multiple sources—wine descriptions from the product catalog, expert tasting notes, and even customer reviews. These texts are preprocessed (cleaning, tokenisation, and possibly translation if the site is multi-lingual) so they can be fed into the classification model.

- Wine Classification Module (HAN + GloVe): The core of our approach is deployed as a microservice or pipeline stage that takes the prepared text and produces classification outputs. These outputs might include the wine's category (variety or region prediction), a flavour profile vector, and key attention-weighted terms that characterise the wine. The hierarchical attention allows this module to effectively summarise each wine in terms of the most salient descriptors.
- Product Metadata Enrichment: The outputs from the classifier are then used to enrich the product database. Each wine's entry is updated with structured tags, for example, flavour: "Berry, Oak, Spice", Body: "Medium" and region: "Napa Valley" (predicted). This enriched metadata becomes part of the wine's profile on the site. Such metadata enrichment improves search and filtering—customers can filter by flavour notes or style, even if those were not originally provided by the merchant, because the AI has inferred them. It also aids content management by maintaining a consistent taxonomy across thousands of product descriptions.
- Recommendation Engine: With both user behaviour data (ratings, past purchases, browsing history) and AI-generated wine profiles, the recommendation engine can generate personalised suggestions. It might use a hybrid of collaborative filtering and content-based filtering. HAN-derived features serve as a content vector for each wine, enabling the engine to find similar wines in the flavour space. For a given user, the system can blend "people who bought X also bought Y" with "our model thinks Y is similar in taste to X" to yield recommendations that are novel yet relevant. The result is a dynamic recommendation module that updates in real-time as new data or new wines come in, always aligning suggestions with the user's demonstrated preferences.
- User Interaction Layer: Finally, the architecture includes the front-end components that interact with customers. This includes personalised recommendation carousels on the homepage ("Recommended for You"), smart search bars that understand natural language queries (if a user types "smooth red wine with chocolate notes", the system uses the classifier's knowledge to return matching wines), and even chatbots or virtual assistants that can answer questions. For instance, a chatbot could be asked, "I need a spicy, full-bodied wine under $30," and because the AI has tagged wines with "spicy" or similar descriptors, it can reliably suggest options. This user layer closes the loop, using the AI's intelligence to engage the customer. The attention mechanism's output could even be used here for explanation: the interface might display a highlighted snippet of the wine's description (generated from the model's attention-weighted words) to show why a particular bottle was recommended, increasing transparency.

Industry adoption ultimately hinges on how the five validated modules can be wired into an e-commerce stack. The architecture that builds on Figure 1 and includes these elements is shown in Figure 6.

The components in Figure 6 form an end-to-end AI-driven ecosystem within the e-commerce site. New wines are analyzed and categorised, and customers continuously benefit through smarter discovery tools. This architecture shines in its adaptability as each component can function as a distinct microservice that can be independently deployed and scaled according to business needs. Existing platforms can implement these modules incrementally through API integration, without requiring whole system replacements. The feedback loops between user interactions and both the recommendation engine and classification model ensure the system continuously evolves with customer behaviour, addressing the cold-start problem that plagues many recommendation systems. This approach mirrors successful AI implementations in retail, where specialised machine learning capabilities are layered onto existing infrastructure to enhance product discovery rather than replacing core systems. By connecting unstructured wine descriptions with structured

searchable attributes, the architecture can provide more relevant recommendations, intuitive natural language search, and consistent product information. As shown in the image evaluation section, it can create a more engaging and personalised shopping experience that drives customer retention and sales growth.

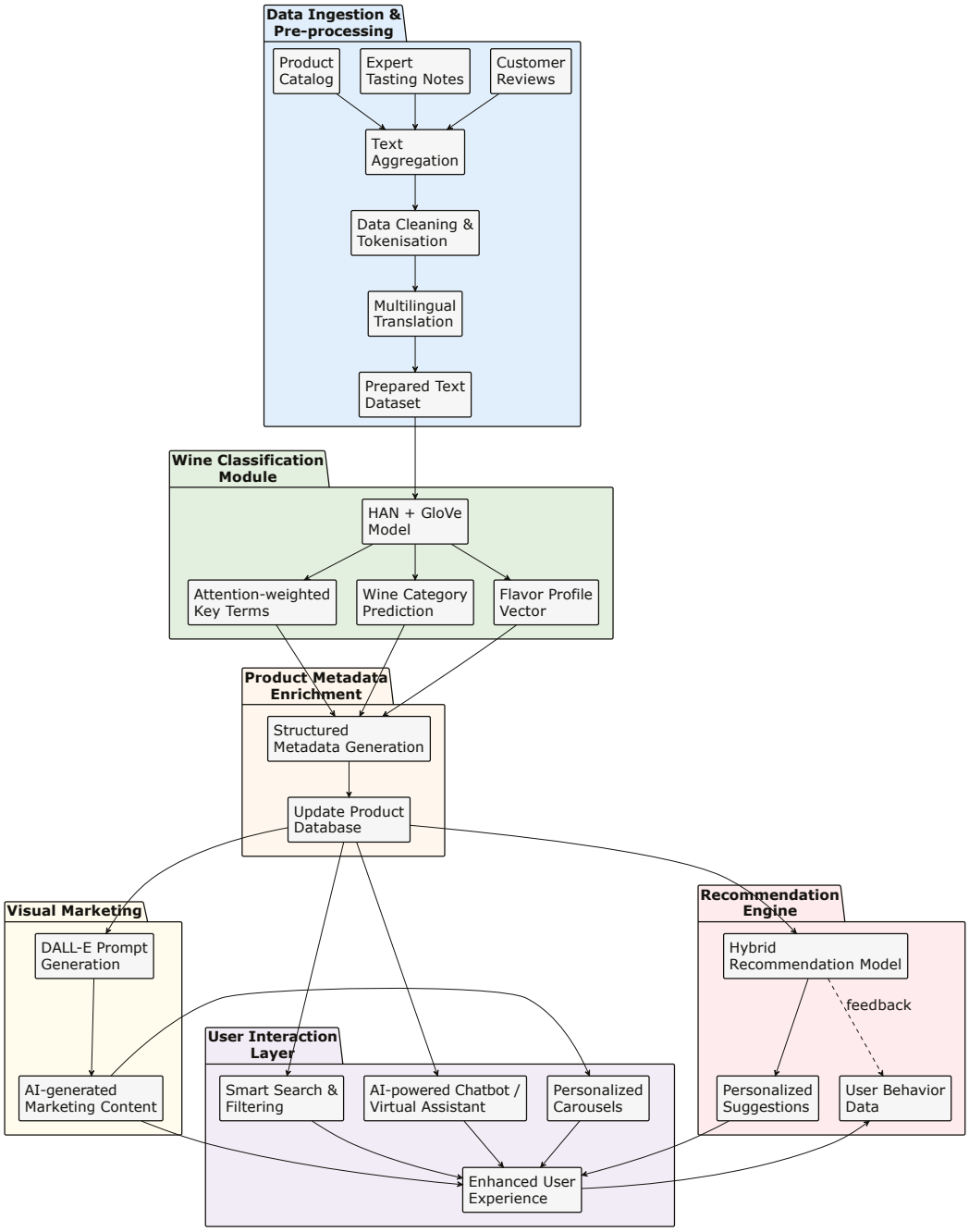


Figure 6. The proposed modular architecture of the system.

5.3. Perception of AI-Generated Marketing Images

Alongside text analytics, we evaluated how buyers react to the visuals that often accompany wine descriptions online. Table 9 makes clear that stylistic coherence matters: when participants perceived a strong visual gap between two AI generators, almost nine in ten leaned toward the more cohesive set. The finding aligns with earlier work on brand consistency and suggests that, in marketing imagery, harmony of style can outweigh small technical differences.

Table 10 adds a second layer. Novice tasters rewarded creative or emotive visuals, whereas professionals penalised lapses in technical polish and brand fit. In practical terms, the same bottle may need different imagery depending on whether the target audience is a casual shopper or a category manager. Coupled with our text classifier, these insights point toward a unified content strategy: prompts and post-processing for AI images should mirror the flavour cues and brand attributes surfaced by the HAN, but they should also be tuned to the expertise level of the intended viewer. The recommendation architecture sketched in Figure 6 can accommodate that logic by selecting or even regenerating images on the fly, ensuring visual and textual consistency across customer touch-points.

5.4. Future Directions

Although our research focused on textual data, there is exciting potential to extend it with multimodal AI techniques in the future. Wine is inherently a multisensory product that includes visuals (label design, colour), taste and aroma (chemical composition, flavour notes), and even context (vineyard, vintage), which all contribute to its identity. Integrating these modalities can take our classification and recommendation system to the next level. One direction is to incorporate visual data: images of wine labels or bottle designs could be analyzed alongside the text. The label can carry information (region indicated by winery name, or style implied by design cues), and a computer vision model could extract those features. For example, a picture of a bottle showing a château and French text might reinforce a prediction that the wine is a Bordeaux. So, an AI that "sees" the wine and "reads" the description would be even more robust in classification, working much like a human who recognises a label and recalls its reputation while reading the tasting notes.

Another clear trend is the blurring of the lines between digital and physical experience. By infusing sensory descriptors and interactive media, online platforms try to replicate the rich experience of in-store shopping. AI comes into play by optimising these features: for example, algorithms might determine which product images or description phrasing yield the best engagement or conversion, and adjust content accordingly (i.e., A/B testing at scale, powered by AI). The emerging paradigm of AI-driven marketing is one in which machines continuously learn from consumer interactions (clicks, conversions, and feedback) and autonomously refine marketing tactics. For industry, this promises improved ROI on marketing spend and agility in responding to consumer trends. For example, if a fashion retailer's AI notices a spike in search queries for "sustainable fabrics" plus positive sentiment around that in reviews, it can prompt the marketing team (or automatically launch) to promote that aspect more heavily, aligning with consumer interest. So, marketing teams will work more closely with data scientists, and KPIs may shift toward those that AI can directly optimise (customer lifetime value, real-time satisfaction scores, etc.). Firms that adapt to these emerging trends will gain a competitive advantage through superior customer engagement. Those who lag may find it hard to meet the personalised and immersive experience standards that consumers will come to expect as AI-driven marketing becomes the norm.

Moreover, consumer feedback data (ratings, reviews in text) can be incorporated to continuously refine the model's understanding of what flavour descriptions align with positive preferences. This approach, which combines text, images, and possibly taste metrics, could lead to a more well-rounded understanding of wine. It opens up novel marketing strategies too: imagine an AR (augmented reality) app where a user scans a wine bottle with their phone and the app (powered by our integrated model) recognises the wine, classifies its profile, and instantly displays an AI-generated tasting summary, maybe even with a virtual sommelier avatar explaining the wine's character. In an online store,

an AI could allow interactive elements, such as showing a flavour wheel graphic for each wine, generated from both text and chemical data, to help users visualise the taste.

These technologies do not operate in isolation. Their integration amplifies their impact. A hierarchical attention model might use GloVe embeddings to analyze which sensory terms in reviews most influence sentiment, combining the first three concepts directly. Likewise, personalisation algorithms might incorporate insights from sentiment analysis (e.g., not recommending a product to a user who left a negative review about it, a form of integrating voice of customer into recommendations). AI-driven marketing is increasingly complex: understanding text, images, and behaviour together to model the entire customer journey. Furthermore, greater use of graph neural networks and knowledge graphs to connect disparate data about consumer behaviour, and an emphasis on ethics and transparency as AI pervades marketing.

We plan to explore training the model on combined datasets (like pairing each wine's description with its label image from a large database such as WineSensed [96], or with structured attributes like region and grape variety) to see if the classifier can learn an even more accurate representation of wines. The goal would be to push the accuracy higher and unlock predictions that are not possible with text alone. For instance, classifying wines by quality rating or how well they age, which might require these additional cues. By extending into multimodal AI, we align our work with the broader trend in machine learning that seeks to mimic human-like understanding: humans use sight, taste, and language together when evaluating wine, so an AI that aspires to assist in wine selection should do the same. This will further strengthen the model's utility in marketing strategies, as campaigns could be tailored based on a more comprehensive profile of each wine (for example, highlighting a bottle's award-badge image along with its top flavour notes in promotional content). The research presented here can set the stage for these advancements, demonstrating the foundational step of deep text understanding, onto which vision and sensory intelligence can be added.

5.5. Cultural and Ethical Considerations

Generative models open a tension between visual novelty and the craft-centric narratives prized in wine culture. AI-rendered labels or personality images may overshoot local iconography, eroding perceptions of authenticity that drive premium pricing. A lightweight mitigation is disclosure: flag AI-generated imagery in the product page meta-data so consumers can decide whether origin matters. More broadly, prompt design should respect appellation codes (for example, banning invented château façades for Bordeaux AOC wines) and avoid visual clichés that flatten regional diversity. Future work should pair user-testing with cultural-heritage scholars to gauge when generative augmentation adds value versus when it dilutes symbolic meaning.

6. Conclusions

We set out to fuse the worlds of wine and words, blending advanced neural architectures with the subtle language of tasting. The result is a Hierarchical Attention Network enhanced by GloVe embeddings that not only captures nuanced flavour notes but also the intangible essence that simpler systems often miss. By converting each label and review into structured searchable content, we have opened new opportunities for online marketplaces to refine how they present, recommend, and ultimately sell wines to curious consumers around the world.

Our text- and image-based approach lays the groundwork for an even more immersive future, a future that encompasses visual cues, chemical signatures, and real-time user feedback to form a "digital tasting" experience rivaling a sommelier's own. It is hardly

far-fetched to imagine a day when a shopper scans a wine bottle with a smartphone and is immediately guided through an AI-driven story of the wine's heritage, flavour profile, best serving suggestions, and evolving peer reviews combined with just the right visual to bring the tasting notes to life.

At its core, our work points out how technology can preserve and amplify the poetry of wine descriptors. By designing a neural architecture with specialised attention layers, we harness the power of word embeddings to read more deeply into tasting notes, capturing context in a manner previously unreachable. As new data streams that flow from label images to augmented reality overlays become more prevalent, we anticipate that models will transcend simple classification, paving the way to personalised recommendations, dynamic content generation, and interactive storytelling.

While observing these possibilities, we get closer to capturing wine's essence in bits and bytes, showing that machine learning pipelines can support both analytic and creative workflows in wine e-commerce. We envision a world in which data-driven sommeliers seamlessly integrate into online retail, bridging the best of tradition with the potential of the digital age and delivering wine experiences that are richer, more intuitive, and delightfully surprising.

Author Contributions: Conceptualisation, V.D.; methodology, V.D. and A.B.; software, V.D.; validation, A.M.I.C., V.D., and I.S.; formal analysis, V.D. and A.M.I.C.; investigation, I.S. and A.B.; resources, V.D.; data curation, V.D.; writing—original draft preparation, V.D., I.S., A.B., and A.M.I.C.; writing—review and editing, V.D., I.S., A.B., and A.M.I.C.; visualisation, V.D., I.S., A.B., and A.M.I.C.; supervision, V.D.; project administration, A.B. and I.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable as the study involved only secondary analysis of publicly available, non-identifiable text (wine reviews) and the evaluation of synthetic generated images.

Informed Consent Statement: Informed consent was obtained electronically before participation. Participants were informed about the purpose of the study, the voluntary nature of participation, procedures and duration, and that only anonymous responses would be used for research.

Data Availability Statement: The wine review dataset constructed and analyzed as part of this study can be downloaded from https://github.com/vladdiac/winereviews (accessed on 13 June 2025). The website constructed for image evaluation can be found at https://github.com/vladdiac/wine-evaluation (accessed on 13 June 2025) and the generated images can be found within the "static" folder.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Examples where generic sentiment tools misread enological terms.

Description Fragment	Compound	Variety	Country
Lemon-soaked nectarine aromas pull the nose into this bottling, which also offers hints of grapefruit and toasted wood.	0.000	Viognier	US
This is a light, crisp and refreshing Sauvignon Blanc. It offers flavours and aromas of green apple, lemon, lime and passion-fruit.	0.000	Sauvignon Blanc	US

Table A1. *Cont.*

Description Fragment	Compound	Variety	Country
Baked cherry and rose-petal aromas are wrapped in oak smoke and tar on the nose of this bottling.	−0.178	Pinot Noir	US
Tightly alkaline aromas of chalk and lime pith meet with crushed sandstone on the nose.	−0.422	Chardonnay	US
Legendary winemaker Randall Grahm breaks ground yet again on this pink blend of Tibouren, Cinsaut and Mourvèdre.	0.000	Rosé	US
Crisp boysenberry and shaved-cedar aromas make for a familiar and likable nose on this bottling.	−0.681	Rhône-style Red Blend	US
Dark and dank green-grass aromas meet with wet concrete and serrano pepper on the nose.	0.000	Sauvignon Blanc	US
Intriguing aromas of crisp apple, green melon, crushed granite and wet river stone make for a zesty nose.	−0.422	Grenache Blanc	US
A smoked-meat aroma overrides fresher notes of blackberry and black plum in this wine.	−0.153	Malbec	US
The wine has structure with tannins and a firm, dry core. It also has ripe acidity and black currant flavours.	0.000	Bordeaux-style Red Blend	France

References

1. Ferrara, M.; Ciano, T. Improving the Interpretability of Asset Pricing Models by Explainable AI: A Machine Learning-based Approach. *Econ. Comput. Econ. Cybern. Stud. Res.* **2024**, *58*, 5–19. [CrossRef]
2. Croijmans, I.; Speed, L.J.; Arshamian, A.; Majid, A. Expertise Shapes Multimodal Imagery for Wine. *Cogn. Sci.* **2020**, *44*, e12842. [CrossRef] [PubMed]
3. Kaimann, D.; Bru, C.L.M.S.; Frick, B. Ratings meet prices: The dynamic relationship of quality signals. *J. Wine Econ.* **2023**, *18*, 226–244. [CrossRef]
4. Rakshit, P.; Sarkar, A. A supervised deep learning-based sentiment analysis by the implementation of Word2Vec and GloVe Embedding techniques. *Multimed. Tools Appl.* **2025**, *84*, 979–1012. [CrossRef]
5. Rizvi, S.M.H.; Imran, R.; Mahmood, A. Text Classification Using Graph Convolutional Networks: A Comprehensive Survey. *ACM Comput. Surv.* **2025**, *57*, 201. [CrossRef]
6. Prakash V, J.; Vijay S, A.A. A multi-aspect framework for explainable sentiment analysis. *Pattern Recognit. Lett.* **2024**, *178*, 122–129. [CrossRef]
7. Hartmann, J.; Exner, Y.; Domdey, S. The power of generative marketing: Can generative AI create superhuman visual marketing content? *Int. J. Res. Mark.* **2025**, *42*, 13–31. [CrossRef]
8. Aragão, M.V.C.; Frigieri, E.P.; Ynoguti, C.A.; Paiva, A.P. Factorial design analysis applied to the performance of SMS anti-spam filtering systems. *Expert Syst. Appl.* **2016**, *64*, 589–604. [CrossRef]
9. Kim, D.; Seo, D.; Cho, S.; Kang, P. Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec. *Inf. Sci.* **2019**, *477*, 15–29. [CrossRef]
10. Onan, A.; Korukoğlu, S.; Bulut, H. Ensemble of keyword extraction methods and classifiers in text classification. *Expert Syst. Appl.* **2016**, *57*, 232–247. [CrossRef]
11. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022. [CrossRef]
12. Castanedo, F.; Aghajan, H.; Kleihorst, R. Modeling and discovering occupancy patterns in sensor networks using latent dirichlet allocation. In *Foundations on Natural and Artificial Computation*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2011; pp. 481–490. [CrossRef]
13. Kim, D.I.; Sudderth, E.B. The doubly correlated nonparametric topic model. In Proceedings of the Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, Granada, Spain, 12–15 December 2011.
14. Gu, D.; Liu, H.; Zhao, H.; Yang, X.; Li, M.; Liang, C. A deep learning and clustering-based topic consistency modeling framework for matching health information supply and demand. *J. Assoc. Inf. Sci. Technol.* **2024**, *75*, 152–166. [CrossRef]

15. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]
16. Liu, B. Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* **2012**, *5*, 1–167. [CrossRef]
17. Nazir, A.; Rao, Y.; Wu, L.; Sun, L. Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey. *IEEE Trans. Affect. Comput.* **2022**, *13*, 845–863. [CrossRef]
18. Pang, B.; Lee, L. A sentimental education. In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Barcelona, Spain, 21–26 July 2004; pp. 271–278. [CrossRef]
19. Maitama, J.Z.; Idris, N.; Abdi, A.; Shuib, L.; Fauzi, R. A systematic review on implicit and explicit aspect extraction in sentiment analysis. *IEEE Access* **2020**, *8*, 170581–170596. [CrossRef]
20. Chifu, A.G.; Fournier, S. Sentiment Difficulty in Aspect-Based Sentiment Analysis. *Mathematics* **2023**, *11*, 4647. [CrossRef]
21. Huang, C.; Han, Z.; Li, M.; Wang, X.; Zhao, W. Sentiment Evolution with Interaction Levels in Blended Learning Environments: Using Learning Analytics and Epistemic Network Analysis. *Australas. J. Educ. Technol.* **2021**, *37*, 81–95. [CrossRef]
22. Katumullage, D.; Yang, C.; Barth, J.; Cao, J. Using Neural Network Models for Wine Review Classification. *J. Wine Econ.* **2022**, *17*, 27–41. [CrossRef]
23. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
24. Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; Hovy, E. Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 1480–1489. [CrossRef]
25. Yang, C.; Cao, J. Interpretable Sentiment Analysis Using the Attention-Based Multiple Instance Classification Model: An Application to Wine Reviews. *Harv. Data Sci. Rev.* **2025**, *7*. [CrossRef]
26. Windsor, L.C.; Tatara, J.H.; Peters, C.B.; Kronsted, C.; Windsor, A. The language of wine reviews. *J. Wine Res.* **2023**, *34*, 81–100. [CrossRef]
27. Pighin, A. Social Media Marketing for Wineries: A Qualitative Analysis of the Adoption of Instagram. Master's Thesis, University of Padova, Padova, Italy, 2023.
28. Bitakou, E.; Karetsos, S.; Ntalianis, F.; Ntaliani, M.; Costopoulou, C. Evaluating Social Media Marketing in the Greek Winery Industry. *Sustainability* **2023**, *16*, 192. [CrossRef]
29. Ferreira, C.; Robertson, J.; Lam, J.; Vella, J. Expert reviews uncorked: Contrasting the differences in the language used in online reviews of white and red wine. *J. Wine Res.* **2023**, *34*, 140–154. [CrossRef]
30. Caballero, R.; Suárez-Toste, E. A genre approach to imagery in winespeak. In *Researching and Applying Metaphor in the Real World*; Low, G., Todd, Z., Deignan, A., Cameron, L., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2010; pp. 265–287. [CrossRef]
31. Kana, M. The Flavour of Words: A Study of Standardised Vocabularies and How Olfactory, Gustatory and Haptic Attributes in Wine Reviews are Currently Rendered in English. Master's Thesis, Utrecht University, Utrecht, The Netherlands, 2023.
32. Bhardwaj, P.; Tiwari, P.; Olejar, K.; Parr, W.; Kulasiri, D. A machine learning application in wine quality prediction. *Mach. Learn. Appl.* **2022**, *8*, 100261. [CrossRef]
33. Clarin, J.A. Comparison of the Performance of Several Regression Algorithms in Predicting the Quality of White Wine in WEKA. *Int. J. Emerg. Technol. Adv. Eng.* **2022**, *12*, 20–26. [CrossRef] [PubMed]
34. Baykal, H.; Yildirim, H.K. Application of Artificial Neural Networks (ANNs) in Wine Technology. *Crit. Rev. Food Sci. Nutr.* **2013**, *53*, 415–421. [CrossRef]
35. Kumar, S.; Kraeva, Y.; Kraleva, R.; Zymbler, M. A Deep Neural Network Approach to Predict the Wine Taste Preferences. In *Modern Management Based on Big Data I and Artificial Intelligence*; Watada, J., Xu, S., Wu, B., Eds.; IOS Press: Amsterdam, The Netherlands, 2020; pp. 1165–1173. [CrossRef]
36. Mittal, K.; Gill, K.S.; Kumar, M.; Rawat, R. Predicting the Red Wine Quality using Exploratory Data Analysis and VGG19 Convolutional Neural Network in Deep Learning. In Proceedings of the 2024 5th International Conference for Emerging Technology (INCET), Belagavi, India, 24–26 May 2024; pp. 1–5. [CrossRef]
37. Basyal, L. *Wine Recommendation System Based on BERT*; ResearchGate Preprint; Stevens Institute of Technology: Hoboken, NJ, USA, 2023. Available online: https://www.researchgate.net/publication/372829666_Wine_Recommendation_System_Based_on_BERT (accessed on 1 August 2025). [CrossRef]
38. Pulluri, K.K.; Kumar, V.N. Wine Quality Assessment using Electronic Nose. In Proceedings of the 2021 Asian Conference on Innovation in Technology (ASIANCON), Melaka, Malaysia, 27–29 August 2021; pp. 1–5. [CrossRef]
39. Yang, Y.; Liu, H.; Gu, Y. A Model Transfer Learning Framework With Back-Propagation Neural Network for Wine and Chinese Liquor Detection by Electronic Nose. *IEEE Access* **2020**, *8*, 105278–105285. [CrossRef]
40. Florea, A.; Sipos, A.; Stoisor, M.C. Applying AI Tools for Modeling, Predicting and Managing the White Wine Fermentation Process. *Fermentation* **2022**, *8*, 137. [CrossRef]

41. Murru, C.; Chimeno-Trinchet, C.; Díaz-García, M.E.; Badía-Laíño, R.; Fernández-González, A. Artificial Neural Network and Attenuated Total Reflectance-Fourier Transform Infrared Spectroscopy to identify the chemical variables related to ripeness and variety classification of grapes for Protected Designation of Origin wine production. *Comput. Electron. Agric.* **2019**, *164*, 104922. [CrossRef]
42. Silva, R.; Freitas, O.; Melo-Pinto, P. Evaluating the generalization ability of deep learning models: An application on sugar content estimation from hyperspectral images of wine grape berries. *Expert Syst. Appl.* **2024**, *250*, 123891. [CrossRef]
43. Rui, M.; Sparacino, A.; Merlino, V.M.; Brun, F.; Massaglia, S.; Blanc, S. Exploring consumer sentiments and opinions in wine E-commerce: A cross-country comparative study. *J. Retail. Consum. Serv.* **2025**, *82*, 104097. [CrossRef]
44. Claster, W.B.; Caughron, M.; Sallis, P.J. Harvesting Consumer Opinion and Wine Knowledge Off the Social Media Grape Vine Utilizing Artificial Neural Networks. In Proceedings of the 2010 Fourth UKSim European Symposium on Computer Modeling and Simulation, Pisa, Italy, 17–19 November 2010; pp. 206–211. [CrossRef]
45. Lefever, E.; Hendrickx, I.; Croijmans, I.; Van den Bosch, A.; Majid, A. Discovering the Language of Wine Reviews: A Text Mining Account. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 7–12 May 2018; pp. 3297–3302.
46. Brand, A.M.M.M.J.; Buica, A. Chemometric and sensometric techniques in enological data analysis. *Crit. Rev. Food Sci. Nutr.* **2023**, *63*, 10995–11009. [CrossRef]
47. Yang, C.; Barth, J.; Katumullage, D.; Cao, J. Wine Review Descriptors as Quality Predictors: Evidence from Language Processing Techniques. *J. Wine Econ.* **2022**, *17*, 64–80. [CrossRef]
48. Chen, B.; Rhodes, C.; Yu, A.; Velchev, V. The Computational Wine Wheel 2.0 and the TriMax Triclustering in Wineinformatics. In *Artificial Intelligence Applications and Innovations*; Springer: Cham, Switzerland, 2016; pp. 223–238. [CrossRef]
49. Dong, Z.; Atkison, T.; Chen, B. Wineinformatics: Using the full power of the computational wine wheel to understand 21st century bordeaux wines from the reviews. *Beverages* **2021**, *7*, 3. [CrossRef]
50. Chen, B. *Wineinformatics: A New Data Science Application*; Springer Nature: Cham, Switzerland, 2022. [CrossRef]
51. Le, L.; Hurtado, P.N.; Lawrence, I.; Tian, Q.; Chen, B. Applying Neural Networks in Wineinformatics with the New Computational Wine Wheel. *Fermentation* **2023**, *9*, 629. [CrossRef]
52. Rong, C.; Zhang, X.; Bi, Y. Food Pairing Based on Generative Adversarial Networks. In *Big Data*; Wang, H., Feng, W., Zhang, Z., Huaming, Y., Bai, J., Guo, Y., Wang, L.M., Eds.; Springer: Singapore, 2021; pp. 148–164. [CrossRef]
53. Park, D.; Kim, K.; Park, Y.; Shin, J.; Kang, J. KitcheNette: Predicting and Ranking Food Ingredient Pairings using Siamese Neural Networks. 2019. Available online: https://github.com/dmis-lab/KitcheNette (accessed on 1 August 2025).
54. Soni, V. Empirical Investigation of the Drivers and Barriers to Adopting Generative AI in Digital Marketing Campaigns. *SSRAMLSageScience* **2023**, *6*, 1–15.
55. McKinsey & Company. How Generative AI can Boost Consumer Marketing. 2023. Available online: https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/how-generative-ai-can-boost-consumer-marketing (accessed on 5 March 2024).
56. Davenport, T.H.; Ronanki, R. Artificial Intelligence for the Real World. *Harvard Business Review* **2018**, *96*, 108–116.
57. Chui, M.; Bughin, J.; Hazan, E.; Ramaswamy, S.; Allas, T.; Dahlström, P.; Henke, N.; Trench, M. *The Economic Potential of Generative AI: The Next Productivity Frontier*; Technical Report; McKinsey Global Institute: New York, NY, USA, 2023.
58. Rivera, K. Between Creatives and Coders: Identifying Potentials of Generative Art in the Creative Processes of Marketing Agencies. Master's Thesis, Zurich University of Applied Sciences School of Management and Law, Zurich, Switzerland, 2023.
59. Ho, J. 100+ Surprising AI Marketing Statistics for 2025. 2025. Available online: https://www.zebracat.ai/post/ai-marketing-statistics (accessed on 22 March 2025).
60. Ramšak, M. Wine Storytelling. In *Slovenian Wine Stories and Anticipations*; Springer: Cham, Switzerland, 2022; pp. 19–21. [CrossRef]
61. Crespo, C.F.; Ferreira, A.G.; Cardoso, R.M. The influence of storytelling on the consumer–brand relationship experience. *J. Mark. Anal.* **2023**, *11*, 41–56. [CrossRef]
62. Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; Chen, M.; Sutskever, I. Zero-shot text-to-image generation. In Proceedings of the 38th International Conference on Machine Learning, Virtual Event, 18–24 July 2021. [CrossRef]
63. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models From Natural Language Supervision. *arXiv* **2021**, arXiv:2103.00020. [CrossRef]
64. Vinetur. How Artificial Intelligence (AI) is Transforming the Wine Industry. 2023. Available online: https://www.vinetur.com (accessed on 10 February 2024).
65. Zhang, X.; Guo, F.; Chen, T.; Pan, L.; Beliakov, G.; Wu, J. A Brief Survey of Machine Learning and Deep Learning Techniques for E-Commerce Research. *J. Theor. Appl. Electron. Commer. Res.* **2023**, *18*, 2188–2216. [CrossRef]
66. Kim, J.M.; Park, K.K.; Marjerison, R.K. Usefulness of Online Reviews of Sensory Experiences: Pre- vs. Post-Pandemic. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 1471–1492. [CrossRef]

67. Li, X.; Li, S.; Li, J.; Yao, J.; Xiao, X. Detection of fake-video uploaders on social media using Naive Bayesian model with social cues. *Sci. Rep.* **2021**, *11*, 16068. [CrossRef]
68. Ratmele, A.; Thakur, R. OpExHAN: Opinion extraction using hierarchical attention network from unstructured reviews. *Soc. Netw. Anal. Min.* **2022**, *12*, 148. [CrossRef]
69. Zhang, W.; Wang, M. An improved deep forest model for prediction of e-commerce consumers' repurchase behavior. *PLoS ONE* **2021**, *16*, e0255906. [CrossRef] [PubMed]
70. Alsubari, S.N.; Deshmukh, S.N.; Aldhyani, T.H.H.; Al Nefaie, A.H.; Alrasheedi, M. Rule-Based Classifiers for Identifying Fake Reviews in E-commerce: A Deep Learning System. In *Handbook of Research on Machine Learning and Security Practices and Applications*; Deshmukh, S.N., Aldhyani, T.H.H., Sharma, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2023; pp. 257–276. [CrossRef]
71. Malik, N.; Bilal, M. Natural language processing for analyzing online customer reviews: A survey, taxonomy, and open research challenges. *PeerJ Comput. Sci.* **2024**, *10*, e2203. [CrossRef] [PubMed]
72. Yıldız, E.; Güngör Şen, C.; Işık, E.E. A Hyper-Personalized Product Recommendation System Focused on Customer Segmentation: An Application in the Fashion Retail Industry. *J. Theor. Appl. Electron. Commer. Res.* **2023**, *18*, 571–596. [CrossRef]
73. Hutto, C.J.; Gilbert, E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, Ann Arbor, MI, USA, 1–4 June 2014; pp. 216–225.
74. Loria, S. TextBlob: Simplified Text Processing. 2024. Available online: https://textblob.readthedocs.io/en/dev/ (accessed on 2 October 2024).
75. Akbik, A.; Bergmann, T.; Blythe, D.; Rasul, K.; Schweter, S.; Vollgraf, R. FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations), Minneapolis, MN, USA, 2–7 June 2019; pp. 54–59. [CrossRef]
76. Anderson, T.; Sarkar, S.; Kelley, R. Analyzing public sentiment on sustainability: A comprehensive review and application of sentiment analysis techniques. *Nat. Lang. Process. J.* **2024**, *8*, 100097. [CrossRef]
77. Bonacchi, C.; Witte, J.; Altaweel, M. Political uses of the ancient past on social media are predominantly negative and extreme. *PLoS ONE* **2024**, *19*, e0308919. [CrossRef] [PubMed]
78. Tymann, K.M.; Steinkamp, L.; Zhurakovskaya, O.; Gips, C. Native Sentiment Analysis Tools vs. Translation Services—Comparing GerVADER and VADER. In Proceedings of the CEUR Workshop Proceedings, Amsterdam, The Netherlands, 1–2 July 2020.
79. Lossio-Ventura, J.A.; Gallivan, D.; Boylston, M.; Tse, S.H.S.; Greenhalgh, T.; Moon, H.; Ji, S.; Tao, C.; Zheng, W.; Zhou, S. A Comparison of ChatGPT and Fine-Tuned Open Pre-Trained Transformers (OPT) Against Widely Used Sentiment Analysis Tools: Sentiment Analysis of COVID-19 Survey Data. *JMIR Ment. Health* **2024**, *11*, e50150. [CrossRef]
80. Mahrukh, R.; Shakil, S.; Malik, A.S. Sentiments analysis of fMRI using automatically generated stimuli labels under naturalistic paradigm. *Sci. Rep.* **2023**, *13*, 7267. [CrossRef]
81. Belay, N. Network and sentiment analysis of Enron emails. *F1000Research* **2024**, *13*, 1485. [CrossRef]
82. Kaur, G.; Kaur, A.; Khurana, M.; Damasevicius, R. Sentiment polarity analysis of love letters: Evaluation of TextBlob, Vader, Flair, and Hugging Face transformer. *Comput. Sci. Inf. Syst.* **2024**, *21*, 1411–1433. [CrossRef]
83. Aljedaani, W.; Babar, M.A.; Mahmood, H.; Biswas, S.; Zhang, Y. Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry. *Knowl.-Based Syst.* **2022**, *255*, 109780. [CrossRef]
84. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral Normalization for Generative Adversarial Networks. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018. [CrossRef]
85. Pennington, J.; Socher, R.; Manning, C. Glove: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543. [CrossRef]
86. Kingma, D.P.; Ba, J.L. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015. [CrossRef]
87. Oppenlaender, J.; Linder, R.; Silvennoinen, J. Prompting AI Art: An Investigation into the Creative Skill of Prompt Engineering. *Int. J. Hum.–Comput. Interact.* **2024**, 1–23. [CrossRef]
88. Katsarou, K.; Douss, N.; Stefanidis, K. REFORMIST: Hierarchical Attention Networks for Multi-Domain Sentiment Classification with Active Learning. In Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, Tallinn, Estonia, 27–31 March 2023. [CrossRef]
89. Cao, J.; Li, X.; Zhang, L. Is Relevancy Everything? A Deep-Learning Approach to Understand the Effect of Image–Text Congruence. *Manag. Sci.* **2023**, *Articles in advance.*. [CrossRef]
90. Hörberg, T.; Kurfalı, M.; Olofsson, J. Chemosensory Vocabulary in Wine, Perfume and Food Product Reviews: Insights from Language Modeling. *Food Qual. Prefer.* **2025**, *124*, 105357. [CrossRef]
91. Gu, T.; Zhao, H.; He, Z.; Li, M.; Ying, D. Integrating External Knowledge into Aspect-Based Sentiment Analysis Using Graph Neural Networks. *Knowl.-Based Syst.* **2023**, *259*, 110025. [CrossRef]

92. Luo, A.; Quadri-Felitti, D.; Mattila, A. The Double-Edged Effects of Visualizing Wine Style: Sweetness Scale on Wine Labels. *Int. J. Contemp. Hosp. Manag.* **2024**, *36*, 2824–2841. [CrossRef]
93. Ilieva, I.; Terziyska, M.; Dimitrova, T. From Words to Ratings: Machine Learning and NLP for Wine Reviews. *Beverages* **2025**, *11*, 80. [CrossRef]
94. Gu, C.; Jia, S.; Lai, J.; Chen, R.; Chang, X. Exploring Consumer Acceptance of AI-Generated Advertisements: Perceived Eeriness and Intelligence. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 108. [CrossRef]
95. Cillo, P.; Rubera, G. Generative AI in Innovation and Marketing Processes: A Roadmap of Research Opportunities. *J. Acad. Mark. Sci.* **2024**, *53*, 684–701. [CrossRef]
96. Bender, T.; Sørensen, S.M.; Kashani, A.; Hjorleifsson, K.E.; Hyldig, G.; Hauberg, S.; Belongie, S.; Warburg, F. Learning to Taste: A Multimodal Wine Dataset. *arXiv* **2023**, arXiv:2308.169002023.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Dynamic Pricing for Internet Service Platforms with Initial Demand Constraints: Unified or Differentiated Pricing?

Junchang Li [1], Jiaqing Sun [2] and Jiantong Zhang [2,*]

[1] School of Economics and Management, Shanghai Polytechnic University, Shanghai 201209, China; lijc@sspu.edu.cn
[2] School of Economics and Management, Tongji University, Shanghai 200092, China; 2010098@tongji.edu.cn
* Correspondence: zhangjiantong@tongji.edu.cn

Abstract

Internet service platforms dynamically charge service prices to satisfy the time-varying service demand by leveraging both full- and part-time service providers. This study developed a dynamic pricing model for a monopolistic service platform under two pricing strategies: unified pricing and differentiated pricing. The model incorporates key factors such as demand fluctuations, initial demand constraints, and service quality. It proved the optimal dynamic pricing scheme aimed at maximizing the platform's expected revenue and analyzed the equilibrium gap between the two strategies based on the optimal control theory. The results reveal the following: (a) The service quality elasticity coefficient, potential market, and demand fluctuation factor all positively affect the optimal service price under the two types of pricing strategies, whereas service quality has the opposite effect. (b) Regardless of pricing strategy, the initial service demand restriction negatively affects the optimal price of the platform. The gap between the optimal service prices under the two types of pricing strategies narrows as the potential service demand rises when customers are less sensitive to service price. (c) With initial demand restriction, the optimal service price rises over time as long as the service market satisfies specific conditions, but the expected revenue under the two types of pricing strategies evolves in significantly different trajectories. (d) The differentiated pricing strategy can help the platform improve revenue by setting a lower revenue-sharing ratio.

Keywords: service platform; time-varying demand; unified pricing; differentiated pricing; demand constraint

1. Introduction

Internet service platforms (abbreviated as service platforms) have emerged as critical infrastructure in the digital economy. Service platformization fundamentally reshapes service delivery across industries ranging from transportation (e.g., ride-hailing, take-out) and healthcare to cloud computing [1,2]. It has been propelled by a confluence of factors, including robust policy support, intense market stimulation, and the booming of the digital economy, especially the catalysis of the COVID-19 pandemic [3]. Service platforms spanning common industries leverage service diversity and demand dynamism to mediate real-time interactions between providers and customers, and generate service value through operational efficiency [4]. Pricing, undoubtedly, is one of the most direct and crucial management tools for service platforms [5]. A reasonable service pricing strategy not only balances supply and demand but also ensures the platform's sustainable operation. Conversely, an unreasonable pricing strategy can severely undermine the platform's

operational profits. Meanwhile, dynamic pricing has been recognized as a pivotal tool for balancing supply–demand mismatches and maximizing platform profitability [6]. Critical challenges persist within the realm of research concentrated on dynamic pricing issues associated with the multifaceted operational characteristics of service platforms, which motivate our research. This study pays attention to three prevalent characteristics inherent in the real-world operation of service platforms: the selection of pricing strategies, the response to service demand fluctuations, and the simultaneous involvement of full-time and part-time service providers.

Unified vs. Differentiated Pricing: In practice, service platforms commonly adopt two main pricing strategies, unified pricing and differentiated pricing, to effectively manage customer markets. Notably, many platforms opt for a differentiated service pricing strategy to better serve segmented markets. For example, ride-sharing platforms offer limousine and express services at different price points to attract distinct passenger segments [7]. Crowdsourcing platforms provide merchant and third-party delivery services with varying prices to target different customer groups. Cloud platforms offer primary and professional cloud services at different prices to segmented client markets, and healthcare platforms provide telephone and graphic consultations at different prices to segmented patient markets. Figure 1a illustrates that the WeDoctor Healthcare Platform simultaneously offers telephone consultation and graphic doctor consultation services to patients at different prices. Telephone (graphic) consultation refers to the healthcare service in which doctors provide consultation services to patients through telephone (graphic)-based communication. This digital consultation not only transcends geographical barriers but also offers patients convenient access to professional medical advice. These platforms tailor their services to satisfy customer needs and differentiate themselves through pricing, defined as a differentiated pricing strategy. However, due to limited customer resources, price and quality competition often exists among the diversified services offered by a single platform. To achieve efficient operation, a small number of service platforms, such as Baidu Map ride-hailing, Hua Xiaozhu Taxi, Indriver, Gett, and Via, have integrated the limousine car drivers and the express car drivers into a flat-rate riding service [8]. As depicted in Figure 1b, T3 Trip offers passengers a flat-rate ride-hailing service at a uniform price. Healthcare platforms like Jingdong Healthcare Platform consolidate all available doctors (high and low professional levels) as a quick consultation service to treat patients with varying symptoms [9]. These platforms charge a uniform price for a unified service by pooling different service providers, defined as a unified pricing strategy. Service platforms with a unified pricing strategy randomly match service providers and customers to attract two-sided users. Each pricing strategy has distinct advantages [9]. Understanding the optimal usage scenarios of these pricing strategies can empower service platforms to make more scientific decisions, enhance their competitiveness, and drive sustainable growth. Despite the prevalence of these two strategies, the specific conditions under which each strategy is most applicable in a service platform remain largely unexplored. This significant knowledge gap forms a core motivation for our research.

Initial Demand Constraints: Customer demand on service platforms is inherently dynamic and fluctuating over time, which is influenced by a multitude of internal and external factors. For example, during peak commute hours, the demand for taxi services skyrockets [10], and during the COVID-19 epidemic, the demand for healthcare services surged due to limited physical medical resources. In response to such demand fluctuations, service platforms typically rely on dynamic pricing as a key mechanism for demand management. Additionally, initial sales volume restrictions and initial inventory limitations [11] play a critical role in the dynamic operation of a retailer. The results of research related to initial demand constraints exert a substantial influence across multiple operational dimensions,

including—but not limited to—strategic pricing, platform design, and revenue management. This practice, often regarded as "Hunger Marketing," aims to stimulate greater demand and increase service prices in the long run by limiting customer access and service scale at the beginning [12]. Many service platforms impose restrictions on customer service demand during their initial operational phase due to the influence of environmental factors such as performance limitations and policy regulations. Cloud computing platforms limit the number of concurrent users and resource requests during working hours to maintain service quality, and healthcare platforms restrict the number of patient appointments and consultations per day to enhance the patient experience. In real-world scenarios, the initial operation phase of service platforms is unique. Whether platforms actively or passively control customer demand during the initial phase, they can guarantee a smooth and stable service commencement while stimulating the potential customer market. Incorporating these initial demand constraints into the dynamic pricing model is essential for accurately reflecting service platform operations, thus further motivating our research to develop a more comprehensive and realistic pricing framework.

Figure 1. Pricing strategies for service platforms: (**a**) WeDoctor platform with differentiated pricing; (**b**) T3 Trip platform with unified pricing.

Most service platforms host both full-time and part-time service providers to engage in service activities. The employment model, characterized by the coexistence of two distinct types of service providers, has found widespread application in ride-hailing platforms [13,14]. On these platforms, full-time drivers typically engage in daily work exceeding 7 h, whereas part-time drivers commit less than 3 h per day. Ride-hailing companies recognize the unique contributions of each driver category and thus implement diverse incentive mechanisms tailored to the two types of drivers [14]. The model extends beyond the ride-hailing industry. For instance, both full-time and part-time nurses play crucial roles in the realm of homecare services [15]. When formulating work schedules for homecare, it is of paramount importance to accommodate the flexibility inherent in part-time nurses' participation [16]. As a flexible employment relationship, part-time workers can buffer the human resource demands of enterprises [16]. It also has a significant impact on individual health [17] and income [18], public management, and social welfare [19,20]. Specifically, part-time employment enables female workers to better balance professional responsibilities with family obligations [21]. On the Didi ride-hailing platform, nearly 90 percent of affiliated drivers engage in part-time driving activities. Similarly, as of September 2024, approximately 73 percent of active physicians affiliated with tertiary hospitals participate in part-time work through the haodf.com Health platform. The part-time pattern illustrates

the increasing flexibility of employment models in the digital era, which reflects a broader transformation in labor market dynamics. Thus, pricing decisions on service platforms are significantly influenced by the composition of registered service providers. Full-time service providers are engaged by service platforms through fixed wages to ensure a stable service supply [22], while part-time providers are attracted by splitting revenue to capture fragmented service value [23]. In this study, we use fixed wages and sharing ratios to represent the participation of both types of service providers, respectively, as their participation profoundly impacts the platform's cost structure and its pricing decisions. This aspect has received comparatively limited attention in the existing literature. It thereby reinforces the motivation for this study to address the gap and contribute to a more comprehensive understanding of service platform pricing.

For service platform operation, the wage of service providers and the investment in service construction constitute the operation costs, while fulfilled service orders generate revenue. Service platforms must make dynamic pricing decisions to strike a balance between revenue and costs under time-varying customer demand, formulated as an optimal control issue. Despite the importance of this issue, research on the optimal dynamic pricing of diversified services remains scarce. Solving the optimal control issue designed to mirror real-world operation is fraught with challenges. To bridge this gap and contribute to the field, we focus on the dynamic pricing problem of a general service platform, adapting a differentiated or unified pricing strategy. Our research aims to answer several key questions: Which pricing strategy should the service platform adopt to maximize expected revenue? How can the service platform dynamically set the service price to adjust supply and demand under initial service demand restrictions? What are the differences in the optimal service prices and the expected revenues under the two types of pricing strategies?

To address these questions, we develop a dynamic pricing model for service platforms under unified and differential pricing strategies, taking into account initial service demand restriction, time-varying customer demand, service quality, and two types of service providers' participation. We then utilize optimal control theory to analyze the characteristics of the optimal dynamic price and revenue from mathematical and numerical perspectives. By regarding service quality as an endogenous variable, the basic dynamic pricing model is further extended to verify the robustness of some key findings. The main contributions of this study are three-fold.

Firstly, this study comprehensively examines the impact of customer demand fluctuation, service competition, initial demand constraints, and pricing strategies. Our findings offer valuable guidance for all kinds of service platforms to optimize their operational revenue through making dynamic pricing decisions. Specifically, we find that, regardless of initial service demand restriction, service quality elasticity coefficient, potential market, and demand fluctuation factor all positively affect the optimal service price, while service quality has a negative impact. Moreover, regardless of the pricing strategy, the initial service demand restriction negatively affects the platform's optimal price. These insights can help service platforms anticipate how changes in crucial factors will influence pricing and make proactive adjustments to their operational strategies.

Secondly, this paper determines the optimal dynamic pricing and revenue for a general service platform. By comparing the differences under the two pricing strategies and investigating the impact of crucial factors, we provide a theoretical foundation and decision support for all kinds of service platforms. For example, we derive that with initial demand restriction, both the optimal unified price and the optimal differentiated price increase over the operation time when there is a higher price elasticity coefficient and the potential customers' scale is below a certain threshold. Moreover, as the potential service demand grows and customers are less price sensitive, the gap between the optimal service

prices under the two pricing strategies narrows. However, when the platform limits initial demand, the expected revenue evolution trajectories under the two pricing strategies differ significantly. These detailed findings not only enrich the theoretical understanding of service platform pricing but also offer practical tools for platforms to make more reasonable pricing decisions, ultimately improving their operational efficiency and profitability.

Thirdly, we enhance the basic dynamic pricing model by regarding service quality as an endogenous variable. Through these enhancements, we further verify the robustness of some key findings. Our analysis reveals that when the synergistic effect of two types of service sharing ratios exceeds a specific threshold, service platforms under both pricing strategies can carry out the optimal pricing schedule to maximize the discount revenue. Moreover, we find that when the service quality elasticity coefficient exceeds a certain threshold, service quality exerts positive and negative effects on the optimal unified service price and the optimal differentiated service price, respectively. These research findings not only strengthen the theoretical foundation of service platform pricing research but also provide a more comprehensive understanding of pricing decisions in dynamic market environments.

The remainder of this paper is laid out as follows. Section 2 reviews the related literature. Section 3 describes the general dynamic pricing problem with associated assumptions. Section 4 analyses the optimal dynamic pricing and revenue of a service platform without or with initial service demand restriction under two types of pricing strategies, respectively. Section 5 carries out a comparative analysis. Section 6 numerically analyzes the combined effect of the operation time and some crucial parameters on the platform's pricing and revenue. Some extensions under endogenous service quality are provided in Section 7. Finally, Section 8 highlights some of the findings and discusses future research.

2. Literature Review

This study investigates the optimal dynamic pricing decisions for a general service platform with diversified service providers under a unified or differentiated pricing strategy. Thus, this section concentrates on three streams of the literature: dynamic pricing for service platforms, pricing strategies of service systems, and pricing for a gig economy service platform.

2.1. Dynamic Pricing for Service Platforms

The dynamic pricing strategy, in contrast to the fixed pricing strategy, empowers service providers to adapt prices with flexibility in response to the continuation of operations, alterations in service status, or fluctuations in market conditions.

Recently, a significant number of scholars and practitioners have continuously demonstrated their interest in the dynamic pricing issue faced by service platforms, but the majority of their research focuses on the pricing of a single ride-hailing service. Cachon et al. (2017) and Lin & Zhou (2019) compared and analyzed the operational impacts of three different pricing strategies—static pricing, dynamic pricing, and surge pricing—on ride-hailing services, which implied that the surge pricing strategy shows more advantages [24,25]. Similarly, many scholars have studied the effectiveness and superiority of surge pricing extensively in the operation of ride-hailing platforms from a theoretical model perspective [26,27]. Nourinejad & Ramezani (2020) developed a dynamic pricing model for a ride-hailing platform and solved it using an optimal controller [10]. Characterizing fluctuations in ridesharing service supply and demand, Chen et al. (2020) proposed a dynamic transaction model for drivers and passengers and presented the optimal dynamic decision-making of a ride-hailing platform [28]. In the research on the optimal pricing of ride-hailing services, Mejjaouli and Tadj (2023) found that the growth rate of the passenger market actively contributes to an increase in ride-hailing prices [29]. The aforementioned

research, however, ignores competition between services, service demand restrictions, and different types of service providers in service platforms.

2.2. Pricing Strategies of Service Systems

Differentiated pricing and unified pricing are the pricing strategies commonly adopted in service systems. Although numerous scholars have conducted studies on the differentiated pricing of service systems (companies), there is a dearth of research on the topic of service platforms' differentiated pricing. Using queuing theory, Maglaras et al. (2018) analyzed the characteristics of optimal pricing and service quality for differentiated services in a single large-scale service system. The results showed that the service system should implement differentiated pricing when segmented markets with low delay sensitivity have sufficient elasticity [6]. Yu et al. (2019) pointed out in their empirical investigation that two competing shipping companies can generate more profits under differentiated pricing than a uniform pricing strategy [30]. Miao et al. (2020) in their research demonstrated that differentiated pricing can not only increase the revenue of electric vehicle rental service companies but also enhance customer satisfaction levels [31]. Ma and Guo (2020) studied the optimal arrival rate and service rate control problems of two types of queues and indicated that, when customers have a relatively low waiting-time cost, the optimal price monotonically increases with the growth of the potential customer scale [32]. Considering the competition between taxis and ride-hailing on the passenger side, Zhao et al. (2022) applied bilateral market theory to study the optimal differentiated pricing of a ride-hailing platform [33]. The above literature concentrated on static differentiated pricing decisions for the service systems; however, it ignores the fluctuation and time-varying features of service demand.

Moreover, uniform pricing has been shown to positively influence customers' purchasing propensity by shaping their value perceptions, while also mitigating the adverse effects commonly associated with differential pricing [34]. For instance, Suning and Gome have adopted unified pricing across both online and offline channels as part of their omnichannel transformation strategies [35]. Empirical evidence from Fan et al. (2023) further supports the efficacy of uniform pricing and demonstrates that unified carbon pricing can reduce regional economic costs [36]. In the context of live-streaming commerce, Wang et al. (2025) conducted a comparative analysis of brand owners' optimal decisions and equilibrium outcomes under uniform and differential pricing strategies from the cannibalization effect [37]. The dual-channel services strategy in the service system is essentially a differential services strategy. Taking two types of services as dual-channel to reach the customer market, our work is also closely related to dual-channel pricing operated by a company. However, the existing research on dual-channel pricing mainly covered fields such as pure dual-channel pricing [35], coordinating [38], and profit mode decision [39]. Considering customers' free-riding effect, Guo et al. (2022) and Li et al. (2019) investigated the dual-channel optimal pricing of a retailer who sells products through both offline and online channels [40,41]. To the best of our knowledge, most dual-channel pricing studies were conducted in the tangible commodity market, rarely involving the service market, especially dynamic service pricing.

2.3. Pricing for a Gig Economy Service Platform

Gig economy service platforms represented by ride-hailing platforms commonly draw in both full-time and part-time service providers. Most existing research on the pricing of ride-hailing platforms has been conducted based on the assumption that only part-time service providers are involved. Some scholars recently have also extended pricing analysis to ride-hailing platforms by incorporating privacy policies [42], car rental services [43],

and autonomous vehicles [44] from the perspective of driver and passenger behavioral utility. For instance, Lin et al. (2024) performed a comparative analysis of optimal service pricing determined by a ride-hailing platform under asset-light and asset-heavy modes [45]. For a gig economy service platform, Chen et al. (2025) analyzed the impact of bounded rationality among customers and service providers on the platform's equilibrium status [46]. Additionally, a growing number of studies have examined how gig economy platforms determine users' access fees [47–49].

For gig economy service platforms, the decentralized (separating/dedicated) mode is associated with differential pricing, whereas the centralized (pooling) mode is linked to unified pricing. As far as we know, fewer research works have explored the pricing of a gig economy service platform through theoretical analysis. Zhong et al. (2020) developed pricing models for an on-demand service platform with two service modes: delivering services to low-sensitivity customers only and to two types of customers simultaneously. They highlighted that the service platform aimed at operational revenue should adopt corresponding service modes according to market status, rather than cover large-scale customers blindly [50]. Considering the safety care of female riders, Tang et al. (2021) studied the optimal service price decision-making problem of a ride-hailing platform under a traditional service mode where all users are pooled together, and a mixed service mode in which female drivers can be chosen, respectively. The results showed that the ride-hailing platform pursuing a win–win status should switch from traditional service mode to a mixed one when the mismatch cost of female riders with safety needs exceeds a certain threshold [51]. Wang and Huang (2022) studied the optimal pricing of a ride-sharing platform simultaneously operating two types of services: express and limousine, where there are entry barriers for limousine [7]. Based on the assumption that a ride-hailing platform can operate low- and high-quality services, Zhong et al. (2023) analyzed the differences in the optimal service price and revenue of the ride-hailing platform under dedicated and pooling service modes and revealed that the ride-hailing platform with pooling service mode can achieve a win–win situation when the opportunity cost of drivers is not very high and the difference in opportunity costs between the two types of drivers is small [8]. Similarly, Li et al. (2025) solved and analyzed the bilateral pricing model for a monopolistic healthcare platform by separating or pooling high and low professional-level doctors [9].

Regarding dynamic pricing for service platforms, although both the literature [10] and our study leverage optimal control theory to construct a dynamic pricing model and analyze the characteristics of an optimal pricing scheme, our study further takes into account the competitive effects between services and the simultaneous participation of two types of service providers. Regarding the pricing of a gig economy service platform, while both the existing literature [8,43] and our study consider the simultaneous engagement of diverse service providers and compare unified pricing with differentiated pricing, they all employ static equilibrium analysis to explore the optimal pricing of ride-hailing platforms, overlooking the dynamic nature of service demands in the real world. These differences precisely highlight the theoretical significance and application value of this study. Table 1 summarizes some of the more relevant research and shows the differences between our study and the others.

Table 1. The differences between our study and some of the more relevant research.

References	Platform	Supply	Demand	Service Competition	Pricing Strategy	Method Adopted	Theory
Zhong et al. [50]	On-demand service platform	Only part-time service providers	Static deterministic	NO	Unified pricing	Equilibrium analysis	Two-sided market theory
Nourinejad & Ramezani [10]	Ride-hailing platform	Only part-time drivers	Dynamic deterministic	NO	Unified pricing	Optimal control	Optimal control theory
Tang et al. [51]	Ride-hailing platform	Male and female drivers	Static deterministic	NO	Unified vs. differentiated pricing	Equilibrium analysis	Two-sided market theory
Zhong et al. [8]	Ride-hailing platform	High-quality and low-quality drivers	Static deterministic	NO	Unified vs. differentiated pricing	Equilibrium analysis	Two-sided market theory
Li et al. [9]	Healthcare platform	Low- and high-level professional doctors	Static deterministic	NO	Unified vs. differentiated pricing	Equilibrium analysis	Two-sided market theory
Our study	Service platform	Full-time and part-time service providers	Dynamic volatility	YES	Unified vs. differentiated pricing	Optimal control	Optimal control theory

Overall, despite the significant research achievements of numerous scholars in the pricing of a gig economy platform, the majority of these studies rely on static equilibrium analysis. They overlook the time-varying and fluctuating nature of service demand and the competitive interactions among services. Our study fills a gap not covered by previous work in two key aspects. Firstly, we abstracted the common characteristics of the real-world operations of service platforms across various fields. These characteristics encompass time-varying demand, competition among services, and the involvement of diversified service providers. Leveraging these abstractions, we constructed a dynamic pricing model for a general service platform. This model is more closely aligned with the actual operations of the platform, and the conclusions drawn from it are more universal.

Secondly, to maximize expected revenue, the optimal pricing scheme of a general service platform was derived by leveraging optimal control theory. Our study comprehensively compares scenarios of no initial demand constraints vs. with initial demand constraints, exogenous vs. endogenous service quality, as well as unified pricing vs. differentiated pricing strategy, contributing to the theoretical framework of dynamic pricing on service platforms. This comprehensive research allows for a more in-depth understanding of how crucial factors like initial demand restriction and service quality influence dynamic pricing decisions under a unified or differentiated pricing strategy.

3. Problem Formulation

This section outlines the foundational assumptions and the research problem that inform the construction of the model framework. For convenience, the notations used in this paper are summarized in Table 2. We utilize the superscript DU, DD to represent dynamic unified pricing and differentiated pricing strategies, respectively.

In a particular service market, a general service platform (e.g., cloud computing platform, ride-hailing platform, and healthcare platform) operates two types of services and dynamically charges service prices by connecting service providers and customers throughout a finite continuous-time horizon, $T \in Q^+$. The operational duration of the service could be as short as a few months (for example, six months or $T = 1/2$) or as long as a few years (for example, two years or $T = 2$). Thus, the operational duration of the service can be a fraction. This assumption is frequently adopted in research related to dynamic pricing [52] (the assumption of the continuity of operation duration).

Table 2. The notations used throughout this paper.

Parameters							
a	The number of potential customers	b	The evolution trend of customer demand	T	Service operation duration	α_i	Price elasticity coefficient of service type i, $i = 1, 2$
α	Price elasticity coefficient	β_i	Quality elasticity coefficient of service type i, $i = 1, 2$	β	Quality elasticity coefficient	η	The competitive effect of service quality
γ	The competitive effect of service price	θ_i	Revenue-sharing ratio of service type i, $i = 1, 2$	ρ_i	Cost coefficient of service type i's quality	f_i	Initial demand limit of service type i, $i = 1, 2$
f	Initial demand limit of the service platform	w	The cost of hiring full-time service providers	φ	Participation coefficient of full-time service providers	δ	Service quality decay coefficient
Control variables							
$p^{DU}(t)$	Service price at time t under DU strategy, $p^{DU}(t) \in \left[\underline{p}, \overline{p}\right]$	$p_i^{DD}(t)$	Price of service type i at time t under DD strategy, $p_i^{DD}(t) \in \left[\underline{p}, \overline{p}\right]$	$S_i(t)$	The participation degree of full-time service providers for service type i, $S_i(t) \in [0, 1]$		
Status variables							
$\dot{D}$(t)	Demand rate of unified service at time t	$\dot{D}_i$(t)	Demand rate of service type i at time t, $i = 1, 2$	$Q(t)$	Demand for unified service at time t	$Q_i(t)$	Demand of service type i at time t, $i = 1, 2$
$q_i(t)$	Quality of service type i at time t, $i = 1, 2$	$\pi^{DU}(t)$	The profit of the service platform at time t under the DU strategy	$\pi^{DD}(t)$	The profit of the service platform at time t under the DD strategy		

Both full-time and part-time service providers are attracted to the service platform as human resources. They are competent in tasks to meet customers' service needs on the service platform, since both full-time and part-time service providers possess professional capabilities. For instance, on the Haodf.com platform, both full-time and part-time doctors are required to acquire medical practice qualifications before providing medical consultation services. It is reasonable to assume that there is no difference in the effectiveness of customers enjoying services provided by part-time or full-time service providers. The service platform provides a fixed salary $w(w > 0)$ for full-time service providers [53]. Meanwhile, the platform shares revenue with part-time service providers involving type i service at a sharing ratio $\theta_i(\theta_i > 0)$ and $\theta_1 \neq \theta_2$ [54], which is determined and announced in advance by the platform in the form of operating rules and remains unchanged during the operation time. The revenue-sharing mechanism for part-time service providers on the service platform is essentially one kind of gig economy (the assumption concerning the resemblance in professional competencies and remuneration forms between the two types of service providers).

The service platform simultaneously operates two types of services $i = 1, 2$ at a price $p_1(t), p_2(t) > 0$ during $t + dt$ a period where $t \in [0, T]$, and the corresponding service quality is q_1, q_2 with $q_1 \neq q_2 > 0$. Service quality represents the extent to which the service platform streamlines its end-to-end service delivery, which is exemplified by measures such as integrating AI-powered assistance systems and refining two-sided matching algorithms. Following the definitions of previous related works [38,39,41], the cost of platform operating service i is a quadratic function of service quality q_i, defined as $\rho_i q_i^2/2$, where $\rho_i(\rho_i > 0)$ is the operation cost coefficient of service i. For customers, the two types of services are similar in terms of function and have a substitution effect to some extent (the assumption of service operation costs based on service quality).

We further assume that the service demand per customer is 1 unit. As demonstrated in previous research [1,23,54], the effective transaction volume on the service platform equals the number of customers who paid for the service because one customer can directly communicate with any service provider assigned by the platform or the service provider

chosen by herself/himself after payment during the $t + dt$ period (the assumption of effective transaction volume on service platforms).

The service platform dynamically adjusts service price(s) to cope with customer market fluctuations. If a service platform implements a differential pricing strategy, it will charge distinct prices to two types of services to satisfy the specific demands of segmented markets, which is indeed the pricing strategy employed by the majority of service platforms. Conversely, when a service platform decides to set an identical price for two types of services, it typically combines them into a unified service. For instance, ride-hailing platforms such as Hua Xiaozhu Taxi and Didi enable both luxury car drivers and express car drivers to jointly provide a flat-rate service. Similarly, on the healthcare 160 service platform, doctors from top-tier hospitals and non-top-tier hospitals are grouped together to offer quick medical consultation services. The operation mode of the monopoly service platform is shown in Figure 2.

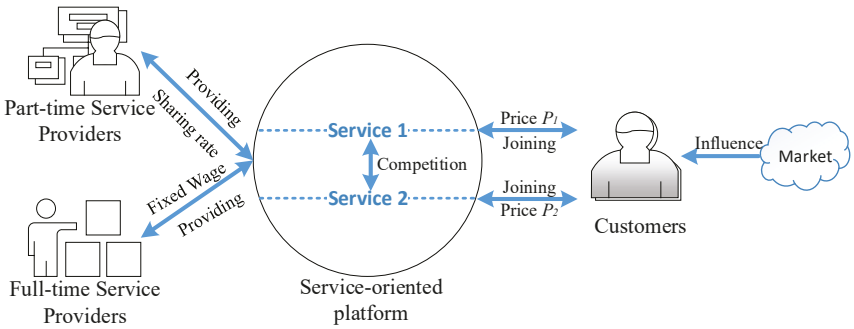


Figure 2. The operation model of a service platform.

Service Demand Rate: Based on the linear demand function [11], the service demand rate function is constructed in this study. Under the influence of the market environment, the service demand evolution of potential customers with size $a(a > 0)$ may exhibit a growth ($b > 0$), decline ($b < 0$), or stability ($b = 0$) trend during the $t + dt$ period. According to the Malthus model, the population growth rate is expressed in an exponential form; thus, the growth rate of potential customers in this study is also expressed in the form of an exponential function, ae^{bt}, which contributes to fluctuations in service demand. According to the customer utility theory, when customers are confronted with a relatively high service price, they will shift away from or withdraw from the current service, thus leading to a decrease in the demand for the service. The negative impact of the service price on the service demand is characterized by the service price elasticity coefficient α_i. When customers perceive a relatively high service quality, they will flock to the service, thereby causing an increase in the service demand. The positive impact of the service quality on the service demand is described by the service quality elasticity coefficient β_i. Simultaneously, grounded in the customer reference theory, when other services with similar functions are priced lower, customers tend to opt for the alternative service with a lower price, which in turn leads to a decline in the demand for the current service. The impact of service price competition on the service demand is characterized by the service price competition coefficient γ. When other services with similar functions have a higher quality, customers tend to choose the alternative service with a higher quality. The impact of service quality competition on the service demand is described by the service quality competition coefficient η. At time t, the customer demand rate of service i operated by the service platform is linearly determined by the potential customer size a, service demand

fluctuation factor b, service price $p_i(t)$, service quality q_i, and competition effect between two types of services [40,53], which is defined as follows:

$$\dot{D}_i(t, p_i(t), p_{3-i}(t)) = ae^{bt} - \alpha_i p_i(t) + \beta_i q_i + \gamma(p_{3-i}(t) - p_i(t)) + \eta(q_i - q_{3-i}),\ i = 1, 2 \quad (1)$$

where α_i, β_i represent the elasticity coefficients of price and quality of service i, respectively. They are, respectively, employed to measure the sensitivity of service demand rate to changes in service price and service quality. γ, η are the competition coefficients of price and quality between the two types of services, respectively, and $0 < \alpha_i, \beta_i, \gamma, \eta < 1$. The service price elasticity coefficient, the service quality elasticity coefficient, and the competitive effect among services are all grounded in factual logic and the practical operation of the service platform.

The service demand evolution is not only affected by current market conditions but also has a carry-over effect to subsequent periods, which reflects the intertemporal characteristics. Let $Q_i(t)$ denote the cumulative demand of service i up to time t, determined by the following:

$$dQ_i(t) = \dot{D}_i(t, p_i(t), p_{3-i}(t))dt \quad (2)$$

Profit Function: Without loss of generality, the other operating costs of the service platform are assumed to be 0 [39,53]. The income of the service platform is closely related to effective service volume, which is defined as $\sum_{i=1}^{2}(1-\theta_i)p_i(t)Q_i(t)$ at time t. The profit function of the service platform at time t is specified as follows:

$$\pi(t, p_i(t), p_{3-i}(t)) = \sum_{i=1}^{2}(1-\theta_i)Q_i(t) - \rho_i q_i^2/2 - w,\ i = 1, 2 \quad (3)$$

where w is considered the fixed operating cost of the service platform, which is utilized to cover the salary expenses for hiring full-time service providers. The total profit of the platform equals revenue minus service operation costs $\rho_i q_i^2/2 + w$.

The service platform initiates its operation at $t = 0$, which is regarded as the initial stage. After operating for a duration of T, the total profit of the service platform can be expressed as follows:

$$\pi(p_i(t), p_{3-i}(t)) = \int_0^T [\sum_{i=1}^{2}((1-\theta_i)dQ_i(t) - \rho_i q_i^2/2) - w]dt,\ i = 1, 2 \quad (4)$$

Dynamic Pricing Model: Dynamic pricing is a strategy that actively adjusts service prices according to real-time market demand. Based on the above description, we formulate a continuous-time dynamic pricing problem for the service platform with diversified services. Let $\boldsymbol{A} = \left\{(p_1(t), p_2(t))|(p_1(t), p_2(t)) \in \left[\underline{p}, \overline{p}\right] \times \left[\underline{p}, \overline{p}\right]\right\}$ denote the platform's admissible control set and $\underline{p}, \overline{p}$ represent the lower and upper bounds of the service price, respectively. The service platform dynamically adjusts the prices of two types of services $(p_1^*(t), p_2^*(t) \in \boldsymbol{A}$ to achieve the maximum profit within the operation time T, defined as

$$\max_{(p_i(t), p_{3-i}(t)) \in \boldsymbol{A}} \pi(p_i(t), p_{3-i}(t)),\ i = 1, 2 \quad (5)$$

and subject to Equation (2).

Specifically, the dynamic pricing model (5) directs the service platform to adopt a unified pricing strategy if $p_1(t) = p_2(t) = p(t)$; otherwise, $p_1(t) \neq p_2(t)$, it adopts a differentiated pricing strategy. Moreover, due to the performance limitations of the service platform in the initial stage, the initial demand should be restricted, $Q_i(0) = f_i > 0$ and

$f = f_1 + f_2$, to the model (5), which can have a cascading effect on subsequent service pricing and demand dynamics. Similarly, initial demand or capacity constraints have been incorporated into relevant dynamic pricing research [55]. The dynamic pricing problem, subject to initial demand constraints, constitutes a special instance of the problem without demand constraints. In reality, when an excessive number of customers initiate service requests concurrently, upon surpassing the capacity threshold defined by the maximum concurrent restriction, a service system will generate notifications like "The maximum concurrent capacity has been reached" or "The connection count exceeds the maximum restriction."

4. Dynamic Pricing for a Service Platform Under Exogenous Service Quality

When the service platform defines service quality in advance, the service quality can be regarded as an exogenous variable. Under exogenous service quality, the section considers the initial service demand without and with restriction, respectively, and proves the optimal dynamic pricing and total profit for a service platform adapting the unified or differentiated pricing strategies.

4.1. Dynamic Unified Pricing (DU) Under Exogenous Service Quality

(1) Without Initial Service Demand Restriction

Under the DU strategy, the service platform organizes both full-time and part-time service providers, coordinating their efforts to present a unified service that is clearly observable within the market landscape. At this point, it is important to note that customer demand does not cease to exist. On the contrary, it undergoes a shift and converges specifically into this unified service offering. The reason is that rejecting the unified service means leaving the service platform, resulting in a perceived utility of 0. This convergence indicates a redirection of customer preferences and needs towards the unified service presented by the platform. The service platform sets a uniform price $p(t)$ for the unified service. Meanwhile, the service platform takes the unified service price as the control variable and the cumulative service demand of customers as the status equation. Based on Equation (1), the customer demand rate for the unified service is as follows:

$$\dot{D}(t, p(t)) = \dot{D}_1(t, p(t)) + \dot{D}_2(t, p(t)) \tag{6}$$

where

$$\dot{D}_1(t, p(t)) = ae^{bt} - \alpha_1 p(t) + \beta_1 q_1 + \eta(q_1 - q_2)$$
$$\dot{D}_2(t, p(t)) = ae^{bt} - \alpha_2 p(t) + \beta_2 q_2 + \eta(q_2 - q_1)$$

The competitive effect no longer exists in the demand rate, represented by Equation (6).

The cumulative demand on the service platform up to time t is determined by the following:

$$dQ(t) = \dot{D}(t, p(t))dt \tag{7}$$

The total profit of the service platform can be determined as follows:

$$\pi^{DU}(p(t)) = \int_0^T \{[(1-\theta_1)dQ_1(t) + (1-\theta_2)dQ_2(t)]p(t) - \rho_1 q_1^2/2 - \rho_2 q_2^2/2 - w\}dt \tag{8}$$

Let $\boldsymbol{A_1} = \left\{ p(t) \middle| p(t) \in \left[\underline{p}, \overline{p}\right] \right\}$ denote the platform's admissible control set under a unified pricing strategy. The service platform dynamically adjusts the unified service price

$p^*(t) \in \boldsymbol{A_1}$ covering the operation time T to maximize total profit. Therefore, the dynamic pricing problem (5) can be transformed into

$$\max_{p(t)\in \boldsymbol{A_1}} \pi^{DU}(p(t)) \tag{9}$$

subject to Equation (7).

For the dynamic pricing problem (9), the Lagrange multiplier λ is introduced to construct the Hamiltonian function as follows:

$$H^{DU} = [(1-\theta_1)(dQ(t) - dQ_2(t)) + (1-\theta_2)dQ_2(t)]p(t) - \rho_1 q_1^2/2 - \rho_2 q_2^2/2 - w + \lambda \dot{D}(t)$$

Lemma 1 (The existence of a unified optimal pricing)**.** *The total profit of the service platform is a concave function of the service price under the unified pricing strategy, and there is an optimal dynamic service price $p^*(t)$.*

Please see Appendix A.1 for the proof of Lemma 1.

Proposition 1 (The expressions of the optimal price and the expected revenue)**.** *Under the unified pricing strategy, the optimal pricing decision and total profit for the service platform are as follows*:

$$p^{DU}(t) = \frac{Ce^{-t}(\alpha_1+\alpha_2) + ae^{bt}(k_1+k_2) + K}{\alpha_1 k_1 - \alpha_2(k-k_2)} \tag{10}$$

$$\begin{aligned}\pi^{DU} = \tfrac{1}{2}\{&-2Tw - \rho_1 T q_2^2 + \tfrac{1}{\alpha_1 k_1 - \alpha_2(k-k_2)^2}[\tfrac{k\alpha_2 a^2(e^{2bT}-1)(k_1+k_2)^2}{b} + \tfrac{4akK\alpha_2(k_1+k_2)(1-e^{bT})}{b(b-1)} \\ &+Cbk_1(e^{(b-1)T}-1)(\alpha_1+\alpha_2)^2 + 2bkK\alpha_2(e^{bT}-1)] + 2k_1CKe^{-T}(e^T-1) \\ &(\alpha_1+\alpha_2)^2 + 2k\alpha_2 TK^2 + 2e^{-T}\sinh(T)(\alpha_1 k_1 + \alpha_2 k_2)C^2(\alpha_1+\alpha_2)^2)\}\end{aligned} \tag{11}$$

where let $k = \theta_1 - \theta_2$, $k_1 = \theta_1 - 1 < 0$, $k_2 = \theta_2 - 1 < 0$, $K = \beta_1 k_1 q_1 + \eta k q_1 - q_2(\beta_2 + \eta\theta_1) + q_2\theta_2(\beta_2+\eta)$. k indicates the difference in revenue-sharing ratios. k_1, k_2 represent the opposite of the revenue ratio on the original two types of services, respectively. C is the freedom coefficient, which is determined by an indefinite integral. K can be regarded as the relative value to the platform. Furthermore, $Ce^{-t}(\alpha_1+\alpha_2)$ represents the basic revenue evolution of the platform determined by its structure. $ae^{bt}(k_1+k_2)$ indicates the operational revenue evolution from serving the customer market. The optimal price under the unified pricing strategy can be simply understood as the ratio of revenue from operating a unified service to the economic capacity, as shown below.

$$p^{DU}(t) = \frac{\overbrace{Ce^{-t}(\alpha_1+\alpha_2) + ae^{bt}(k_1+k_2) + K}^{\text{Revenue from operating a unified service}}}{\underbrace{\alpha_1 k_1 - \alpha_2(k-k_2)}_{\text{Economic capacity}}}$$

Please see Appendix A.2 for the proof of Proposition 1.

Theorem 1 (The evolutionary trend of the optimal unified price)**.** *When $C, b > 0$ and $max\left\{0, \frac{\alpha_2(k-k_2)}{k_1}\right\} < \alpha_1 < 1$, noted as Condition 1, the optimal unified price for the service platform increases over time; otherwise, the optimal unified price for the service platform decreases over time. Condition 1 means that customers are sensitive to service price and have a growing service demand.*

Please see Appendix A.3 for the proof of Theorem 1.

Theorem 1 indicates that even if customers are sensitive to service price, the service platform still raises its unified service price over time to maximize total profit in the context that there is a surging trend in customer demand without initial service demand restriction.

Corollary 1 (The impacts of platform performance parameters)**.** *When $0 < \theta_2 < \theta_1$, the elasticity coefficients of service quality (β_1, β_2), alternative service quality (q_2), potential demand (a), and demand fluctuation factor (b) all contribute to an increase in the optimal unified pricing. The optimal unified pricing reduces with an increase in service quality (q_1) when the elasticity coefficient of service quality (β_1) is less than a certain threshold. Under the unified pricing strategy, a negative relationship exists between the total profit of the platform and the cost coefficients of the two types of services (ρ_1, ρ_2) and the fixed salary (w).*

Please see Appendix A.4 for the proof of Corollary 1.

Corollary 1 shows that the quality of alternative services, the elasticity coefficient of service quality, and potential demand positively stimulate customer service demand when the service platform adopts the unified pricing strategy. Driven by huge service demand, the service platform tends to set a higher unified service price (consistent with intuition). However, service quality negatively affects the unified service price when the elasticity coefficient of service quality is smaller. This is because the impact of service quality on service demand is relatively small. Hence, improving service quality does not significantly promote service demand but increases the platform operation costs.

(2) With Initial Service Demand Restriction

Proposition 2 (The expressions of the optimal price and the expected revenue)**.** *When the service platform restricts the initial service demand, $Q(0) = f$, the optimal unified price and total profit of the platform are as follows*:

$$p^{DU}(t) = \frac{Ce^{-t}(\alpha_1 + \alpha_2) + ae^{bt}(k_1 + k_2) + K}{\alpha_1 k_1 - \alpha_2(k - k_2)}$$

$$\pi^{DU} = \int_0^T [((1-\theta_1)(dQ(t) - dQ_2(t)) + (1-\theta_2)dQ_2(t))p^{DU}(t) - \frac{\rho_1 q_1^2}{2} - \frac{\rho_2 q_2^2}{2} - w]dt$$

where $C = \frac{k[a(\alpha_1 - 3\alpha_2) + \alpha_1\beta_2 q_2 + \eta(q_2 - q_1)(\alpha_1 + \alpha_2)] + f[\alpha_1 - \alpha_1\theta_1 + \alpha_2(k - k_2)] - k(2\beta_1 q_1 \alpha_2 + \beta_2 q_2 \alpha_2)}{(\alpha_1 + \alpha_2)^2}$.

The expression of the unified price with initial service demand restriction is the same as that of the unified price without initial service demand restriction, but the freedom coefficient is determined.

Please see Appendix A.5 for the proof of Proposition 2.

Theorem 2 (The evolutionary trend of the optimal unified price)**.** *Considering the initial service demand constraint, the optimal unified price and its evolution intensity steadily rise over time when $3\alpha_2 < \alpha_1 < min\{\frac{\alpha_2(k-k_2)}{k_1}, 1\}$ and $0 < a < \frac{f[\alpha_1 k_1 - \alpha_2(k-k_2)] + A}{(\alpha_1 - 3\alpha_2)k + b^2 B}$, noted as Condition 2, and under the unified pricing strategy; otherwise, the optimal unified price and its evolution intensity decline over time. Additionally, the initial service demand restriction hurts the optimal unified price of the service platform. Condition 2 means that the customers' sensitivity to service price and the number of potential customers are, respectively, less than a certain threshold.*

Please see Appendix A.6 for the proof of Theorem 2.

Theorem 2 reveals that in the case of an initial service demand constraint, the service platform gradually raises the unified service price as the operation continues when

customers exhibit relatively low sensitivity to service price and the potential customer demand is less than a certain threshold (consistent with intuition). At the same time, the increase in the unified service price is more significant in the later stages of operation due to the cumulative effect of service demand. With the initial service demand restriction, service demand and service price rise as the platform continues to operate. This section explains a common market phenomenon that most service companies prefer "hunger marketing" regardless of the market demand trend, such as crowd-sourcing platforms, cloud computing platforms, and healthcare platforms.

Corollary 2 (The evolutionary trend of the expected revenue)**.** *Considering the initial service demand constraint, the total profit of the service platform rises with the increase in operating times when* $0 < f < max\{|F_1|, |F_2|\}$*, noted as Condition 3, where* F_1, F_2 *are the real roots of* $\frac{\partial \pi^{DU}(T)}{\partial T} = 0$ *on f. Otherwise, the total profit of the service platform declines over operation time. Condition 3 means that the platform limits the initial service demand to less than a certain threshold.*

Please see Appendix A.7 for the proof of Corollary 2. Corollary 2 suggests that if the service platform adopts the unified pricing strategy, it is better to keep the initial service demand below a certain level to achieve sustained profit, which provides theoretical support for the unified pricing decision-making of service platforms.

In conclusion, regardless of whether or not the initial service demand is constrained, the optimal unified price of a service platform increases over time when certain conditions hold.

4.2. Dynamic Differentiated Pricing (DD) Under Exogenous Service Quality

(1) Without Initial Service Demand Restriction

Under the DD strategy, the service platform simultaneously organizes full-time and part-time service providers to offer two types of services that can be observed by the market, each catering to different market segments. The service platform sets different prices, $p_1(t) \neq p_2(t)$, for two types of services. Meanwhile, the platform takes different service prices as the control variables, $\boldsymbol{p}^{DD}(t) = \{p_1(t), p_2(t)\}$, and the cumulative demand on two types of services as the status equations, $\boldsymbol{Q}(t) = \{Q_1(t), Q_2(t)\}$, where the customer demand rates for service type 1 and 2 are expressed as Equation (1).

The service platform adjusts different service prices covering the operation time T to maximize total profit, which is defined as follows:

$$\pi^{DD}(p_1(t), p_2(t)) = \int_0^T [(1-\theta_1)dQ_1(t)p_1(t) + (1-\theta_2)dQ_2(t)p_2(t) - \frac{\rho_1 q_1^2 + \rho_2 q_2^2}{2} - w]dt \tag{12}$$

Let $\boldsymbol{A_2} = \{(p_1(t), p_2(t)) | (p_1(t), p_2(t)) \in \left[\underline{p}, \overline{p}\right] \times \left[\underline{p}, \overline{p}\right]\}$ denote the platform's admissible control set under a unified pricing strategy. The service platform dynamically adjusts the differentiated service prices $\left(p_1^*, p_2^*\right) \in \boldsymbol{A_2}$ covering the operation time T to maximize total profit. Therefore, the dynamic pricing problem (5) can be transformed into

$$\max_{(p_1(t), p_2(t)) \in \boldsymbol{A_2}} \pi^{DD}(p_1(t), p_2(t)) \tag{13}$$

subject to Equation (2).

For the dynamic pricing problem (9), Lagrange multipliers $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2\}$ are introduced to construct the Hamiltonian function as follows:

$$H^{DD} = (1-\theta_1)dQ_1(t)p_1(t) + (1-\theta_2)dQ_2(t)p_2(t) + \lambda_1 \dot{D}_1(t) + \lambda_2 \dot{D}_2(t) - \frac{\rho_1 q_1^2 + \rho_2 q_2^2}{2} - w$$

Lemma 2 (The existence of a differentiated optimal pricing)**.** *The total profit of the service platform is a concave function of service price under the differentiated pricing strategy, and there is an optimal dynamic service price* $\left(p_1^*, p_2^*\right)$ *when* $\alpha_1 > \frac{\gamma\left[\gamma k^2+4\alpha_2(\theta_1+\theta_2-\theta_1\theta_2-1)\right]}{4k_1k_2(\alpha_2+\gamma)}$*, noted as Condition 4, which means that customers are more sensitive to service price; otherwise, there is no optimal dynamic service price.*

Please see Appendix B.1 for the proof of Lemma 2.

Lemma 2 indicates that the service platform can charge the optimal differentiated price to optimize total profit when the service price elasticity coefficient exceeds a certain threshold. The threshold is determined by the sharing ratio of two types of services, the intensity of price competition, and the elasticity coefficient of the alternative service's price.

Proposition 3 (The expressions of the optimal price and the expected revenue)**.** *Under the differentiated pricing strategy, the optimal pricing decisions and total profit for the service platform are determined as follows*:

$$\begin{aligned} p_1^{DD}(t) = \tfrac{e^{-t}}{(\alpha_2\gamma+\alpha_1R_2)Z}\{&-C_1[\alpha_1R_2+\gamma(\alpha_2+\gamma\theta_1)]+\gamma kC_2R_2+ \\ &aZe^{t+bt}(\alpha_2+2\gamma)+e^tZ[\beta_1q_1R_2+\eta\alpha_2q_1+q_2(\gamma\beta_2-\eta\alpha_2)]\}+C_1\theta_2R_1R_2 \end{aligned} \tag{14}$$

$$\begin{aligned} p_2^{DD}(t) = \tfrac{e^{-t}}{(\alpha_2\gamma+\alpha_1R_2)Z}\{&C_2[\alpha_1k_1R_2+\gamma(\alpha_2k_1+\gamma k)]-\gamma kC_1R_1+ \\ &aZe^{t+bt}(\alpha_1+2\gamma)+e^tZ[\gamma\beta_1q_1+\beta_2q_2R_1+\eta\alpha_1(q_2-q_1)]\} \end{aligned} \tag{15}$$

$$\begin{aligned} \pi^{DD}(T) = &\tfrac{a(C_1+C_2)[e^{(b-1)T}-1]}{b-1}+\tfrac{e^{-T}\sinh(T)[k_1R_2C_2^2+k_2R_1C_1^2-\gamma C_1C_2(k_1+k_2)]}{Z}-\tfrac{\rho_1Tq_1^2+\rho_2Tq_2^2}{2} \\ &+[C_1q_1(\beta_1+\eta)+C_2q_2(\beta_2+\eta)-\eta C_2q_1-\eta C_1q_2][1-\cosh(T)-\sinh(T)]-Tw \end{aligned} \tag{16}$$

where $z = k_1k_2 > 0$, $R_1 = \alpha_1+\gamma$, $R_2 = \alpha_2+\gamma$, *and* $0 < R_1, R_2 < 2$. C_1, C_2 *Nare the freedom coefficients, which are determined by an indefinite integral.* z *represents the synergy effect of services operated by the platform on the revenue rate.* R_1, R_2*, respectively, indicate the elasticity of service demand to prices of service 1 and 2 under the service substitution effect. Interestingly, although there is a similar structure to the optimal unified price, the expression of the optimal service price under the differentiated pricing strategy is more complex as a result of competition and the synergy of alternative services.*

Please see Appendix B.2 for the proof of Proposition 3.

Theorem 3 (The evolutionary trend of the optimal differentiated price)**.** *The optimal differentiated prices of the service platform decrease over time when* $b > 0$ *and* $0 < a < min\{\frac{e^{-(b+1)t}\left[\gamma kC_2R_2+C_1\theta_2R_1R_2-C_1\left(\alpha_1R_2+\gamma\alpha_2+\theta_1\gamma^2\right)\right]}{bZ(\alpha_2+2\gamma)}, 1\}$*, noted as Condition 5, which means the potential service demand from customers is smaller but has an increasing trend; otherwise, the optimal differentiated prices of the service platform rise over time.*

Please see Appendix B.3 for the proof of Theorem 3.

Theorem 3 shows that even though the demand market has a surging trend, the optimal service price of the service platform adopting the differentiated pricing strategy gradually declines when potential customer demand is below a certain threshold, which means the pricing decision of the service platform is significantly influenced by the potential demand market.

Corollary 3 (The impacts of platform performance parameters)**.** *The optimal service price (taking type 1 service as an example) rises with its service quality* (q_1)*, the elasticity coefficients of service quality* (β_1, β_2)*, potential demand* (a)*, and demand fluctuation factor* (b) *when the platform*

adopts a differentiated pricing strategy. When the elasticity coefficient of alternative service quality (β_2) *is less than a certain threshold, the optimal service price decreases with the improvement of alternative service quality* (q_2)*. When there is a quality difference between two types of services* $(q_1 < q_2)$*, the intensity of service quality competition* (η) *positively affects the optimal service price. The cost coefficients of the two types of services* (ρ_1, ρ_2) *and fixed salaries* (w) *harm the total profit of the platform.*

Please see Appendix B.4 for the proof of Corollary 3.

From Corollary 3, it can be seen that, similar to the case of a unified pricing strategy, service quality, elasticity coefficient of service quality, and potential demand market all positively affect the optimal differentiated price. The difference is that alternative service quality positively affects the optimal differentiated price when the elasticity coefficient of alternative service quality is lower. The substitutability between services positively affects the optimal differentiated price when the service quality is smaller. The main reason is that the positive effect of alternative service synergy outweighs the negative effect of competition.

(2) With Initial Service Demand Restriction

Proposition 4 (The expressions of the optimal price and the expected revenue)**.** *When the service platform restricts the initial service demand,* $\boldsymbol{Q}(\boldsymbol{0}) = \{f_1, f_2\} > \boldsymbol{0}$*, the optimal differentiated price and total profit of the platform are derived as follows*:

$$p_1^{DD}(t) = \frac{e^{-t}[ae^{(1+b)t}(\alpha_2 + 2\gamma) + \gamma\beta_2 q_2 + \beta_1 q_1 R_2 + \alpha_2\eta(q_1 - q_2) - \gamma f_2 - f_1 R_2]}{\gamma\alpha_2 + \alpha_1 R_2}$$

$$p_2^{DD}(t) = \frac{e^{-t}[ae^{(1+b)t}(\alpha_1 + 2\gamma) + \gamma\beta_1 q_1 + \beta_2 q_2 R_1 + \alpha_1\eta(q_2 - q_1) - \gamma f_1 - f_2 R_1]}{\gamma\alpha_2 + \alpha_1 R_2}$$

$$\begin{aligned}\pi^{DD} = \frac{1}{2(\gamma\alpha_2+\alpha_1\beta_2)}\{&\frac{2a[e^{(b-1)T}-1][f_1k_1(\alpha_2+2\gamma)+f_2k_2(\alpha_1+2\gamma)]}{1-b} + e^{-2T}[k_1(e^{2T}-1)f_1^2 - 2e^T f_1(e^T-1)(q_1\beta_1R_2 + \eta q_1\alpha_2 + \\ &q_2(\gamma\beta_2 - \eta\alpha_2)]k_1 + k_2R_1f_2^2(e^{2T}-1) + f_1(2e^T(\gamma\beta_1q_1 + \beta_2q_2R_1 + \eta\alpha_1(q_2-q_1))k_2 - \gamma f_1(k_1+k_2) + e^{2T}(2q_1k_2(\eta\alpha_1 - \gamma\beta_1) \\ &-2q_2(\gamma\beta_2 + \alpha_1k_2(\beta_2+\eta) + \gamma f_1(k_1+k_2))) - e^{2T}T(\gamma\alpha_2+\alpha_1R_2)(2w + \rho_1q_1^2 + \rho_2q_2^2))\}\end{aligned}$$

The form of the optimal differentiated price with service demand restriction is similar to that of the optimal unified price with service demand restriction, but all service quality, initial service demand, and service competition factors play an important role.

Please see Appendix B.5 for the proof of Proposition 4.

Theorem 4 (The evolutionary trend of the optimal differentiated price)**.** *Considering the initial service demand constraints, the optimal differentiated price rises over time when the service demand market surges under the differentiated pricing strategy. Additionally, the initial service demand has a negative impact on the optimal differentiated price of the service platform.*

Please see Appendix B.6 for the proof of Theorem 4.

Theorem 4 reveals that if there is a surging trend in customer service demand, implementing "hunger marketing" and adopting the differentiated pricing strategy is beneficial for gradually increasing the optimal differentiated service prices set by the service platform (consistent with intuition), which is consistent with common sense for service platforms operating with a differentiated pricing strategy in fact.

Corollary 4 (The optimal operation duration)**.** *Considering the initial service demand constraints, there is $T^* \in [0, T]$ that makes the service platform achieve the maximum total profit when $b < 1$ and*

$$0 < a < \frac{e^{-(1+b)T} f_1 \{2f_1 R_2 - e^T[\alpha_2\beta_1 q_1 + \gamma(\beta_1 q_1 + \beta_2 q_2) + \eta\alpha_2(q_1 - q_2)] + 2k_2 R_1 f_2^2 + 2\gamma f_1 f_2(k_1 + k_2)\} + e^{-bT} f_2 k_2[\gamma\beta_1 q_1 + \beta_2 q_2 R_2 + \alpha_1\eta(q_2 - q_1)]}{(b-1)[f_1 k_1(\alpha_2 + 2\gamma) + f_2 k_2(\alpha_1 + 2\gamma)]},$$

noted as Condition 6, under the differentiated pricing strategy. Condition 6 means that the service demand of customers with a smaller potential scale has a slowly increasing trend.

Please see Appendix B.7 for the proof of Corollary 4.

Corollary 4 shows that the service platform adopting "hunger marketing" can maximize its total profit within the constrained operational period T^* when the potential service demand and fluctuation factor are lower under the differentiated pricing strategy. Under such circumstances, it is best for the service platform to make a suitable plan for the operation cycle according to the service demand and market characteristics.

In conclusion, under the differentiated pricing strategy, the optimal differentiated price of the service platform decreases over time if the initial service demand is unlimited and the service demand of customers with smaller potential market rises. On the contrary, the optimal differentiated price increases over time if the initial service demand is constrained and shows a surging trend.

5. Comparison Analysis Under Exogenous Service Quality

We theoretically compare the differences in service prices and the differences in operational profit between the two pricing strategies: unified pricing and differential pricing.

Corollary 5 (Service price comparison)**.** *Taking the optimal differentiated price of service 1 as an example, the gap between the optimal unified price and the optimal differentiated price of the service platform narrows as the potential service demand or demand fluctuation factor increases when $0 < \alpha_1 < min\{\frac{\alpha_2(k-k_2)}{k_1}, 1\}$, noted as Condition 7, which means that the elasticity coefficient of service 1 price is lower; otherwise, the gap between the optimal unified price and the optimal differentiated price of the service platform expands as the potential service demand or demand fluctuation factor increases.*

Please see Appendix C.1 for the proof of Corollary 5.

Corollary 5 indicates that when customers' price sensitivity is less than a certain threshold, the gap between the optimal unified price and the optimal differentiated price of the service platform narrows as potential customer demand grows. The partial reason is that service prices adjusted by the service platform primarily follow the market-pulling effect in response to demand situations, but weaken the pricing strategies' role.

Corollary 6 (Profit comparison)**.** *The total profit gap of the service platform between the unified pricing strategy and the differentiated pricing strategy widens with the increasing intensity of service price competition when $C_1 > \frac{k_1 C_2}{k_2}$, $0 < \theta_1, \theta_2 < 1$, noted as Condition 8, which means that the revenue ratio of service 2 is greater than that of service 1; otherwise, the total profit gap of the service platform between the unified pricing strategy and the differentiated pricing strategy narrows with the increasing intensity of service price competition.*

Please see Appendix C.2 for the proof of Corollary 6.

Corollary 6 suggests that strengthening competition among service prices can highlight the differences and advantages of pricing strategies on operational profit when the revenue ratio of one service to part-time service providers is relatively smaller.

6. Numerical Analysis

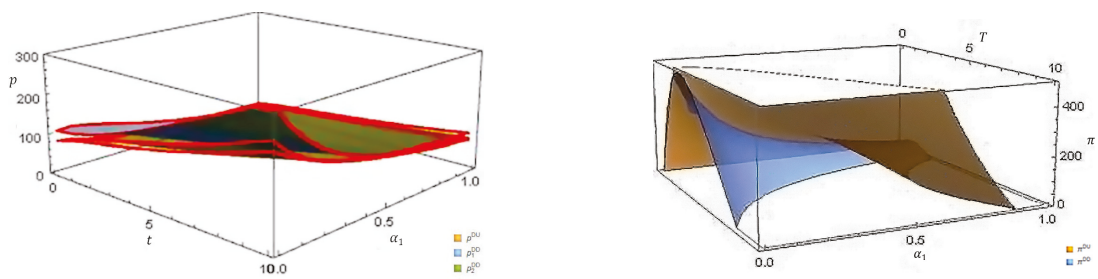
In this section, we utilize numerical calculations to analyze the sensitivity of the service platform's optimal service price and total profit to important factors under two types of pricing strategies. Crucial parameters in this study can be quantified utilizing questionnaires, interviews, and empirical analysis. Taking Jingdong Health as an example, the number of potential patients accessed is greater in reality. The critical parameter values are standardized and displayed in Table 3 according to the actual situation of this healthcare platform.

Table 3. Parameter value settings.

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
a	20	b	0.1	T	10	α_1	0.4
α_2	0.5	β_1	0.7	β_2	0.65	η	0.8
γ	0.75	q_1	7.8	q_2	8.1	θ_1	0.3
θ_2	0.25	ρ_1	1.0	ρ_2	1.2	w	38
f	10	f_1	5	f_2	5	\	\

(1) Service price elasticity

Taking the price elasticity of service 1 as an example, Figure 3a,b, respectively, show the joint impact of the service price elasticity and operation time on the optimal service price and total profit of the service platform under the two types of pricing strategies. From Figure 3a, the service platform prefers to set higher service prices over time due to the cumulative effect of service demand, which balances excess service demand for increasing profitability. It further validates Theorems 1, 2, 3, and 4. As customers become more sensitive to service prices, the optimal service price decreases regardless of the pricing strategy adopted by the service platform. When the elasticity of the service price is small, the evolution trend in the service price is particularly evident under the differentiated pricing strategy; conversely, when the elasticity of the service price is larger, the evolution trend in the alternative service price is particularly apparent under the differentiated pricing strategy. The simulation further validates Corollaries 2 and 4; in this parameter setting, whether the service platform chooses the differentiated pricing strategy or the unified pricing strategy, its profit exhibits, respectively, an approximate inverted "U" shape and a positive relationship with operational time. Additionally, the profit of the service platform decreases as customer sensitivity to service price increases, as Figure 3b shows.



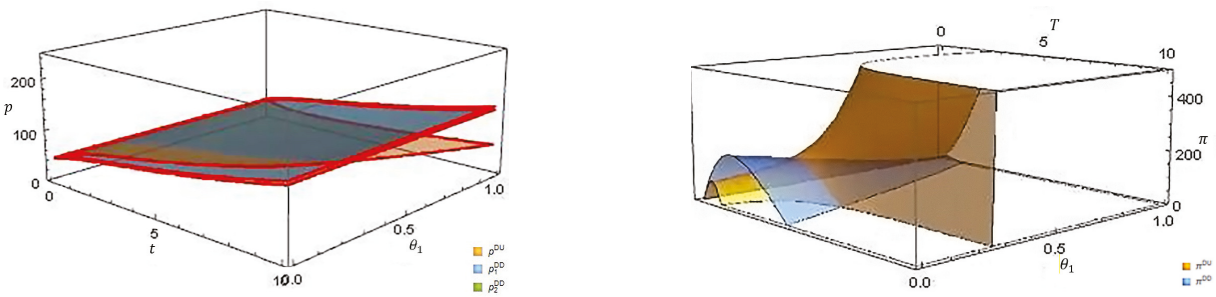
(**a**) (**b**)

Figure 3. The sensitivity of service price and total profit to price elasticity: (**a**) sensitivity of service price to price elasticity; (**b**) sensitivity of total profit to price elasticity.

(2) Revenue-sharing ratio

Taking the revenue-sharing ratio of service 1 as an example, Figure 4a,b, respectively, depict the joint impact of the revenue-sharing ratio and operation time on the optimal service price and total profit under the unified and differentiated pricing strategies. Figure 4a

shows that, in our parameter setting, the evolution trend in optimal service price under two types of pricing strategies remains unchanged when the revenue-sharing ratio is smaller, similar to Figure 3a. In contrast, the growth trend in the optimal service price under the unified pricing strategy is vital. However, the service platform with the unified pricing strategy tends to set a lower service price when the revenue-sharing ratio is larger. As the revenue-sharing ratio rises, the optimal service price under the differentiated pricing strategy gradually grows, while the optimal service price under the unified pricing strategy slowly drops. As Figure 4b shows, the revenue-sharing ratio has the opposite effect on the service platform's profit under two types of pricing strategies. There is an approximate inverted "U" shape relationship between the revenue-sharing ratio and the service platform's profit. Compared to the unified pricing strategy, the higher revenue-sharing ratio under the differentiated pricing strategy means that higher compensation is paid to part-time service providers, which results in the platform bearing higher operating costs. Thus, the service platform should charge a lower sharing ratio when using a differentiated pricing strategy to maximize profit.

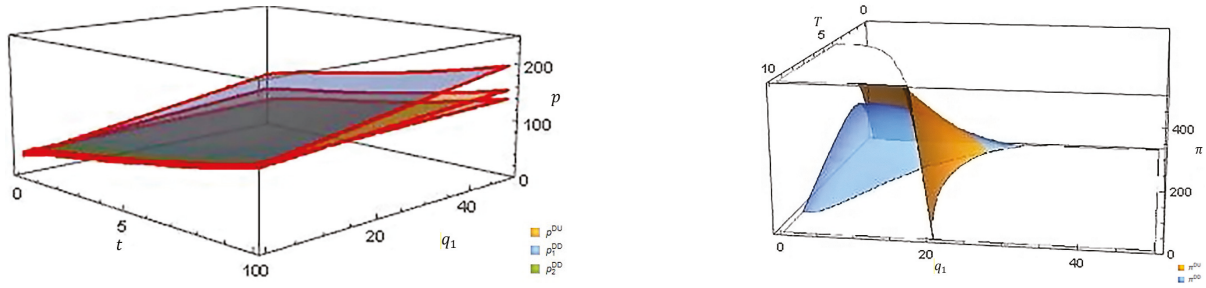


(**a**) (**b**)

Figure 4. The sensitivity of service price and total profit to the revenue-sharing ratio: (**a**) sensitivity of service price to revenue-sharing ratio; (**b**) sensitivity of total profit to revenue-sharing ratio.

(3) Service quality

Taking quality of service 1 as an example, Figure 5a,b, respectively, display the joint impact of service quality and operation time on the optimal service price and total profit under the unified and differentiated pricing strategies. Figure 5a reveals that, similar to Figure 3a, the evolution trend in the optimal service price under the two types of pricing strategies remains unchanged, whereas the growth trend in the optimal service price is greater under the differentiated pricing strategy. Simultaneously, there is a positive relationship between the service quality and the optimal service price under two types of pricing strategies. This is because the service platform provides customers with higher quality service at the cost of higher service investment and operating costs. The optimal revenue of the service platform under two types of pricing strategies decreases accordingly with an improvement of service quality and also has an approximate inverted "U" shape with operational time, as Figure 5b shows.



(**a**) (**b**)

Figure 5. The sensitivity of service price and total profit to service quality: (**a**) sensitivity of service price to service quality; (**b**) sensitivity of total profit to service quality.

7. Extension: Under Endogenous Service Quality

To incorporate the complex relationship between service quality and the engagement of full-time service providers into this model framework, service quality is treated as an endogenous variable. This modelling makes an extension of the basic dynamic pricing structure. In the section, the engagement of full-time service providers, service quality, and revenue discount rate are introduced to improve the original dynamic pricing model and further verify the robustness of the conclusions.

Since the available service time of full-time service providers is relatively stable and can be controlled by the service platform, a reasonable assumption is made: the service quality of type-i service is influenced by the engagement of full-time service providers $s_i(t)$. According to the extended model of Nerlove–Arrow [56], the evolution of service quality for type-i service at the period $t + \Delta t$ is defined as follows:

$$q_i(t + \Delta t) = \varphi s_i(t) - \delta q_i(t),\ q_i(T) = m_i > 0,\ i = 1, 2 \tag{17}$$

where the engagement of full-time service providers is positively correlated with the salary provided by the service platform. φ represents the coefficient of full-time service providers' engagement. δ is the decay rate. And $\varphi,\ \delta > 0$.

7.1. Dynamic Unified Pricing (DU) Under Endogenous Service Quality

From the basic model in Section 3, the functional form of the service demand evolution remains unchanged. The variable branches are further integrated, $\alpha_1 = \alpha_2 = \alpha$, $\beta_1 = \beta_2 = \beta$, and $0 < \alpha, \beta < 1$. The change rate of service demand under the dynamic unified pricing strategy is modified as follows:

$$\dot{D}_i(t, p(t), q_i(t), q_{3-i}(t)) = ae^{bt} - \alpha p(t) + \beta q_i(t) + \eta[q_i(t) - q_{3-i}(t)],\ i = 1, 2 \tag{18}$$

The cumulative demand on the service platform up to time t is determined by the following:

$$dQ(t) = \left[\dot{D}_1(t, p(t), q_1(t), q_2(t)) + \dot{D}_2(t, p(t), q_1(t), q_2(t))\right]dt \tag{19}$$

$$Q(0) = f > 0$$

Let $\boldsymbol{B_1} = \left\{(p(t), s_1(t), s_2(t)) | (p(t), s_1(t), s_2(t)) \in \left[\underline{p}, \overline{p}\right] \times [0, 1] \times [0, 1]\right\}$ denote the platform's admissible control set. The platform should make decisions on both the unified price and the participation level of full-time service providers simultaneously $\left(p^*(t), s_1^*(t), s_2^*(t)\right) \in \boldsymbol{B_1}$ to maximize expected revenue covering the operation time T,

$$\pi^{DU}(p(t),s_1(t),s_2(t)) = \max_{(p(t),s_1(t),s_2(t))\in B_1} \int_0^T e^{-\sigma t}[(1-\theta_1)p(t)dQ_1(t) + (1-\theta_1)p(t)dQ_2(t) - \rho_i s_i^2(t)/2]dt \quad (20)$$

subject to Equations (17) and (19). $e^{-\sigma t}$ and σ are the discount factor and discount rate, respectively.

For the dynamic pricing problem (20), Lagrange multipliers $\lambda_1, \lambda_2, \lambda_3$ are introduced to construct the Hamiltonian function as follows:

$$H^{DU} = (1-\theta_1)p(t)dQ_1(t) + (1-\theta_1)p(t)dQ_2(t) + \lambda_1[\varphi s_1(t) - \delta q_1(t)] + \lambda_2[\varphi s_2(t) - \delta q_2(t)] + \lambda_3(\dot{D}_1 + \dot{D}_2) - \frac{\rho_1 s_1^2(t) + \rho_2 s_2^2(t)}{2}$$

Proposition 5 (The existence of a unified optimal pricing)**.** *Under the unified pricing strategy, there is an optimal decision-making scheme for the service platform.*

$$p^{DU}(t) = \frac{\beta k_1 q_1(t) - q_2(t)(\beta + \eta\theta_1) + \eta k q_1(t) + R_3\theta_2 q_2(t) + ae^{bt}(\theta_1 + \theta_2 - 2)}{2\alpha(\theta_1 + \theta_2 - 2)}$$

$$s_1^{DU}(t) = \frac{\varphi\lambda_1}{\rho_1},\ s_2^{DU}(t) = \frac{\varphi\lambda_2}{\rho_2}$$

where

$$q_1(t) = \frac{{}^{-t\delta - tR_4}p\left(1 - {}^{t\delta + tR_4}\right)(\beta k_1 + \eta k)\varphi^2}{\rho_1 R_4(2\delta + \sigma)} - \frac{{}^{-t\delta}p(\beta k_1 + \eta k)\left({R_4}^{t\delta} + \delta^{-tR_2}\right)\varphi^2}{\delta\rho_2 R_4(2\delta + \sigma)} + {}^{-t\delta}C_1 + \frac{{}^{-t\delta}\left({}^{t\delta + tR_2} - 1\right)\varphi^2 C_2}{\rho_1(2\delta + \sigma)}$$

$$+q_2(t) = \frac{{}^{-t\delta - tR_4}p\left(1 - {}^{t\delta + tR_4}\right)(\beta k_2 - \eta k)\varphi^2}{\rho_2 R_4(2\delta + \sigma)} - \frac{{}^{-t\delta}p(\beta k_2 - \eta k)\left({R_2}^{t\delta} + \delta^{-tR_4}\right)\varphi^2}{\delta\rho_2 R_4(2\delta + \sigma)} + {}^{-t\delta}C_3 + \frac{{}^{-t\delta}\left({}^{t\delta + tR_2} - 1\right)\varphi^2 C_4}{\rho_2(2\delta + \sigma)}$$

$$\lambda_1 = -\frac{p(t)(\beta k_1 + \eta k)}{R_4} + {}^{tR_4}C_2,\ \lambda_2 = -\frac{p(t)(\beta k_2 - \eta k)}{R_4} + {}^{tR_4}C_4,\ \lambda_3 = \frac{pk_1}{1-\sigma} + {}^{t(\sigma-1)}C_5$$

$$R_3 = \beta + \eta,\ R_4 = \delta + \sigma$$

Based on all boundary conditions, the coefficients $C_1 \sim C_5$ can be determined, and the final form of the optimal decisions and expected revenue can be derived.

Please see Appendix D.1 for the proof of Proposition 5.

Corollary 7. *Under the unified pricing strategy, the optimal unified service price increases with the expansion of operating duration* (t)*, service quality elasticity coefficient* (β)*, the number of potential customers* (a)*, and demand volatility factor* (b)*, but decreases with the service price elasticity coefficient* (α) *rising. When the service quality elasticity coefficient exceeds a certain threshold* $(\beta > \{\frac{-\eta k}{k_1}, \frac{\eta k}{k_2}\})$*, the improvement of service quality* (q_1, q_2) *can promote the optimal unified price. When the current service quality is lower than the alternative service quality, the revenue-sharing ratios of current service* (θ_1) *and alternative service* (θ_2) *have opposite effects on the optimal unified price, which are positive and negative, respectively.*

Corollary 7 reveals that, in service platform operations under the unified pricing strategy, the optimal unified service price is intricately shaped by multiple factors. Longer

operating durations, higher service quality elasticity coefficients, larger numbers of potential customers, and increased demand volatility all contribute to price increases. Conversely, a rise in the service price elasticity coefficient, as in ride-hailing platforms where customers are highly price sensitive, leads to price decreases. Moreover, when the service quality elasticity coefficient crosses a certain threshold, such as in a software-as-a-service platform, service quality enhancements can drive up the optimal price (consistent with intuition). These are also similar to some of the findings in Theorems 1 and 2, Corollary 1. Please see Appendix D.2 for the proof of Corollary 7.

7.2. Dynamic Differentiated Pricing (DD) Under Endogenous Service Quality

Similarly, the customer demand rate under the dynamic differentiated pricing strategy is modified as follows:

$$\dot{D}_i(t, p_i(t), p_{3-i}(t), q_i(t), q_{3-i}(t)) = ae^{bt} - \alpha p(t) + \beta q_i(t) + \gamma[p_{3-i}(t) - p_i(t)] + \eta[q_i(t) - q_{3-i}(t)],\ i = 1,\ 2 \tag{21}$$

The cumulative demand on the service i up to time t determined by the following:

$$dQ_i(t) = \dot{D}_i(t, p_i(t), p_{3-i}(t), q_i(t), q_{3-i}(t))dt \tag{22}$$

$$Q_i(0) = f_i > 0$$

Let $\boldsymbol{B_2} = \{(p_i(t), p_{3-i}(t), q_i(t), q_{3-i}(t)) | (p_i(t), p_{3-i}(t), q_i(t), q_{3-i}(t)) \in [\underline{p}, \overline{p}] \times [\underline{p}, \overline{p}] \times [0,1] \times [0,1]\}$ denote the platform's admissible control set. The platform should make decisions on both the differentiated price and the participation level of full-time service providers simultaneously $(p_1^*(t), p_2^*(t), s_1^*(t), s_2^*(t)) \in \boldsymbol{B_2}$ to maximize expected revenue covering the operation time T,

$$\pi^{DD}(p_1(t), p_2(t), s_1(t), s_2(t)) = \max_{(p_1(t), p_2(t), s_1(t), s_2(t)) \in B_2} \int_0^T e^{-\sigma t}[\sum_{i=1}^{2}(1-\theta_i)p_i(t)dQ_i(t) - \rho_i s_i^2(t)/2]dt \tag{23}$$

subject to Equation (22).

For the dynamic pricing problem (23), Lagrange multipliers $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are introduced to construct the Hamiltonian function as follows:

$$H^{DD} = (1-\theta_1)p_1(t)dQ_1(t) + (1-\theta_2)p_2(t)dQ_2(t) + \lambda_1[\varphi s_1(t) - \delta q_1(t)] + \lambda_2[\varphi s_2(t).-\delta q_2(t)] + \lambda_3\dot{D}_1(t) + \lambda_4\dot{D}_2(t) - \frac{\rho_1 s_1^2(t) + \rho_2 s_2^2(t)}{2}$$

Proposition 6 (The existence of a differentiated optimal pricing)**.** *Under the differentiated pricing strategy, there is an optimal decision-making scheme for the service platform when* $z > max\{\frac{\gamma^2 k^2}{4\alpha^2 + 8\alpha\gamma}, \frac{\gamma^2(\theta_1+\theta_2-2)^2}{4R^2}\}$.

$$p_1^{DD}(t) = \frac{k_2\left[q_1(t)(2\beta\gamma k_1 + 2\alpha k_1 R_3 + \gamma\eta k) + q_2(t)(2\alpha\eta - 2\alpha\eta\theta_1 - \gamma\eta\theta_1 + \gamma\eta\theta_2 + \beta\gamma\theta_1 + \beta\gamma\theta_2 - 2\beta\gamma) + ae^{bt}(2\alpha k_1 + 3\gamma\theta_1 + \gamma\theta_2 - 4\gamma)\right]}{4z\alpha^2 + 8\alpha\gamma z - \gamma^2 k^2}$$

$$p_2^{DD}(t) = \frac{k_1\left[q_2(t)(2\beta\gamma k_2 + 2\alpha k_2 R_3 - \gamma\eta k) + q_1(t)(\beta\gamma\theta_1 + \beta\gamma\theta_2 - 2\beta\gamma + 2\alpha\eta + \gamma\eta\theta_1 - 2\alpha\eta\theta_2 - \gamma\eta\theta_2) + ae^{bt}(2\alpha k_2 + \gamma\theta_1 + 3\gamma\theta_2 - 4\gamma)\right]}{4z\alpha^2 + 8\alpha\gamma z - \gamma^2 k^2}$$

$$s_1^{DD}(t) = \frac{\varphi\lambda_1}{\rho_1},\ s_2^{DD}(t) = \frac{\varphi\lambda_2}{\rho_2}$$

where

$$q_1(t) = \frac{-t\delta - tR_4}{\rho_1 R_4(2\delta+\sigma)}\left(1 - {}^{t\delta+tR_4}\right)[k_1 R_3 p_1(t) - \eta k_2 p_2(t)]\varphi^2 - \frac{-t\delta[k_1 R_3 p_1(t) - \eta k_2 p_2(t)]\left(R_4{}^{t\delta} + \delta^{-tR_2}\right)\varphi^2}{\delta\rho_1 R_4(2\delta+\sigma)} + {}^{-t\delta}C_1 + \frac{-t\delta\left(t\delta+tR_4 - 1\right)\varphi^2 C_2}{\rho_1(2\delta+\sigma)}$$

$$q_2(t) = \frac{^{-t\delta - tR_4}\left(1 - {}^{t\delta + tR_4}\right)[k_2R_3p_2(t) - \eta k_1p_1(t)]\varphi^2}{\rho_2R_4(2\delta+\sigma)} + \frac{^{-t\delta}(\eta k_1p_1(t) - k_2R_3p_2(t))\left(R_4{}^{t\delta} + \delta^{-tR_2}\right)\varphi^2}{\delta\rho_2R_4(2\delta+\sigma)} + {}^{-t\delta}C_3 + \frac{^{-t\delta}\left(^{t\delta+tR_4} - 1\right)\varphi^2C_4}{\rho_2(2\delta+\sigma)}$$

$$\lambda_1 = \frac{\eta k_2p_2(t) - R_3k_1p_1(t)}{R_2} + {}^{tR_2}C_2,\ \lambda_2 = \frac{\eta k_1p_1(t) - R_3k_2p_2(t)}{R_2} + {}^{tR_2}C_4,\ \lambda_3 = \frac{p_1(t)k_1}{1-\sigma} + {}^{t(\sigma-1)}C_5,\ \lambda_4 = \frac{p_2(t)k_2}{1-\sigma} + {}^{t(\sigma-1)}C_6$$

$$R = \alpha + \gamma$$

It derives the final forms of the optimal decisions and expected revenue based on all boundary conditions.

Please see Appendix D.3 for the proof of Proposition 6.

Corollary 8. *As the operating duration (t), potential customer scale (a), and demand volatility factor increase (b), the optimal service price rises accordingly. When the service quality elasticity coefficient exceeds a certain threshold ($\beta > max\{\frac{\eta(-2\alpha k_1 - \gamma k)}{2k_1R}, \frac{\eta(2\alpha k_1 + \gamma K)}{\gamma(\theta_1+\theta_2-2)}\}$), both the current service quality (q_1) and the alternative service quality (q_2) have a negative impact on the optimal service price. When the synergistic effect of two types of service sharing ratio exceeds a specific threshold ($\frac{\gamma^2k^2+4\alpha^2(\theta_1+\theta_2-\theta_1\theta_2-1)}{8\alpha\gamma} < z$), the optimal service price increases with the growth of the service quality elasticity coefficient (β).*

Corollary 8 suggests that, under a differential pricing strategy, the effect of service quality on the optimal price is opposite to that under a unified pricing strategy. At this time, the negative effect of competition among services becomes apparent (consistent with intuition). The service price elasticity coefficient exhibits a non-linear effect on the optimal price. Additionally, some of the properties of the findings in Corollary 8 are consistent with those in Corollary 3 and Theorem 4. Please see Appendix D.4 for the proof of Corollary 8.

7.3. Comparison Analysis Under Endogenous Service Quality

Since it is not possible to directly derive the explicit expression for the platform's optimal service price and expected revenue, we use numerical analysis to compare the optimal service prices and the expected revenues between the two pricing strategies, respectively. Let $\alpha = 0.4$, $\beta = 0.6$, $\varphi = 1$, $\delta = 0.8$, $\sigma = 0.7$, $m_1 = m_2 = 0.6$. Other parameter values remain the same.

(1) Service price comparison

As shown in Figure 6, the optimal differential price and the optimal unified price exhibit similar changing trends under the impact of crucial parameters. With the continuous operation of the platform (Figure 6a) or a gradual increase in the demand fluctuation factor (Figure 6d), the optimal service price increases nearly linearly. However, the elasticity coefficient of price (Figure 6b) has a negative effect on the optimal service price. The optimal service price demonstrates an approximately U-shaped relationship with the elasticity coefficient of quality (Figure 6c) or the sharing ratio (Figure 6e,f). Interestingly, when the sharing ratio exceeds a certain threshold, the service platform sharply raises the service price to balance the revenue-sharing cost. In most cases, the price of type 1 service is the highest, and that of type 2 service is the lowest under the differentiated pricing strategy.

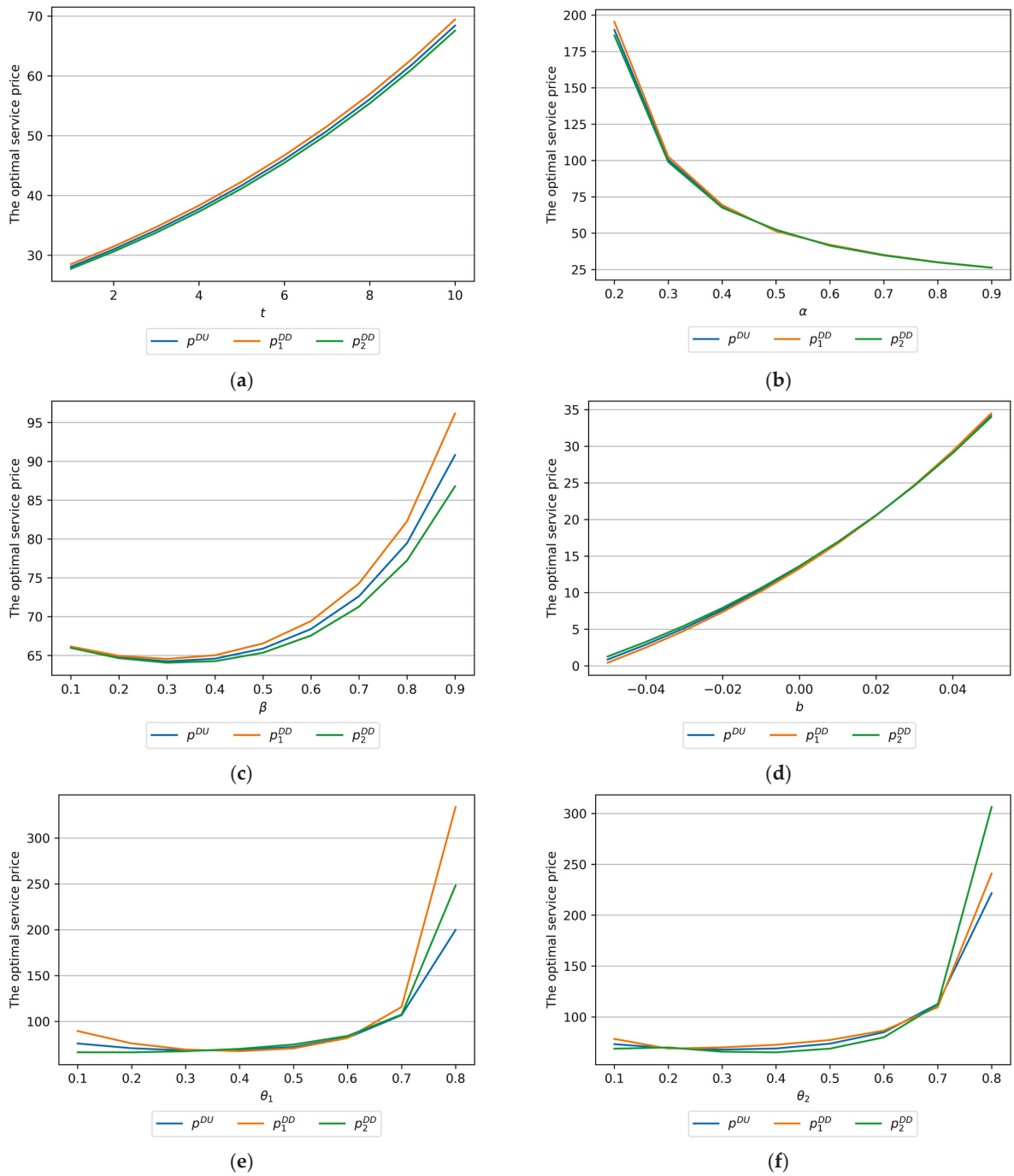


Figure 6. Service price comparison between the two service pricing strategies: (**a**) the impact of operational time; (**b**) the impact of the elasticity coefficient of price; (**c**) the impact of the elasticity coefficient of quality; (**d**) the impact of the demand fluctuation factor; (**e**) the impact of the sharing ratio of service 1; and (**f**) the impact of the sharing ratio of service 2 on the optimal service price.

(2) Revenue comparison

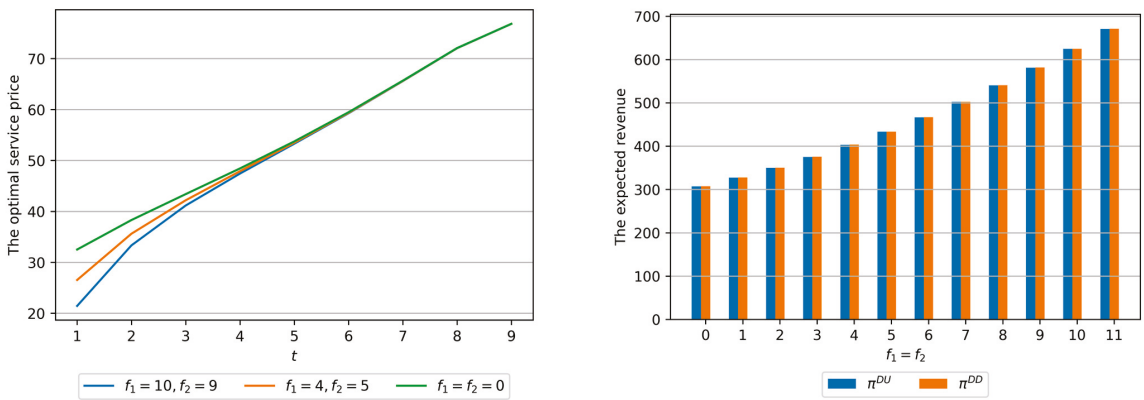
Under the parameter setting, the effects of these crucial parameters on the platform's expected revenue are generally similar to their effects on the optimal service price, as shown in Table 4. In most cases, the differentiated pricing strategy yields a higher expected revenue, which also highlights the advantages of differentiated pricing. However, when the elasticity coefficient of service is greater than a certain threshold, the platform can achieve a higher expected revenue by adopting a unified pricing strategy.

Table 4. Revenue comparison between the two service pricing strategies.

T	π^{DU}	π^{DD}	$\pi^{DU}-\pi^{DD}$	α	π^{DU}	π^{DD}	$\pi^{DU}-\pi^{DD}$	β	π^{DU}	π^{DD}	$\pi^{DU}-\pi^{DD}$
1	123.00	123.09	−0.09	0.1	43394.70	46865.80	−3471.10	0.1	478.28	478.23	0.05
2	260.50	260.70	−0.20	0.2	1478.20	1486.10	−7.90	0.2	448.63	448.42	0.22
3	338.91	339.14	−0.24	0.3	675.98	677.09	−1.10	0.3	428.62	428.64	−0.02
4	384.65	384.94	−0.29	0.4	433.53	433.87	−0.34	0.4	418.57	418.49	0.08
5	409.09	409.41	−0.32	0.5	318.22	318.19	0.03	0.5	419.08	419.24	−0.16
6	421.61	421.94	−0.33	0.6	251.12	251.14	−0.02	0.6	433.53	433.87	−0.34
7	427.90	428.24	−0.34	0.7	207.29	207.28	0.01	0.7	467.58	468.78	−1.20
8	431.05	431.39	−0.34	0.8	176.45	176.43	0.03	0.8	534.30	538.20	−3.90
9	432.61	432.98	−0.37	0.9	153.58	153.55	0.03	0.9	665.27	678.32	−13.05
10	433.53	433.87	−0.34								
b	π^{DU}	π^{DD}	$\pi^{DU}-\pi^{DD}$	θ_1	π^{DU}	π^{DD}	$\pi^{DU}-\pi^{DD}$	θ_2	π^{DU}	π^{DD}	$\pi^{DU}-\pi^{DD}$
−0.05	240.74	240.76	−0.01	0.1	580.45	692.52	−112.07	0.1	536.80	556.55	−19.75
−0.04	249.47	249.50	−0.02	0.2	488.41	508.69	−20.27	0.2	459.89	459.95	−0.05
−0.03	258.63	258.66	−0.03	0.3	433.53	433.73	−0.20	0.3	413.68	417.42	−3.74
−0.02	268.24	268.28	−0.04	0.4	407.37	409.18	−1.81	0.4	392.78	406.60	−13.82
−0.01	278.34	278.39	−0.05	0.5	412.52	418.96	−6.44	0.5	401.42	422.94	−21.52
0	288.95	289.16	−0.20	0.6	468.32	472.76	−4.44	0.6	460.63	480.73	−20.10
0.01	300.12	300.34	−0.21	0.7	665.94	649.14	16.81	0.7	667.07	659.38	7.70
0.02	311.89	312.11	−0.22	0.8	1821.15	2273.98	−452.83	0.8	1959.63	2190.83	−231.20
0.03	324.15	324.53	−0.38								

(3) The impacts of initial demand restriction

The impacts of initial demand constraints on the optimal service price and expected revenue are shown in Figure 7a,b, respectively. In Figure 7a, taking the price of service type 1 as an example, the initial demand constraints are sequentially set as $f_1 = f_2 = 0$, $f_1 = 4$ and $f_2 = 5$, $f_1 = 10$ and $f_2 = 9$. In Figure 7b, the initial demand constraints are sequentially set as $f_1 = f_2 = 0, 1, 2, \cdots, 11$. Under this parameter setting, the initial demand constraint has a negative impact on the service price and expected revenue, which further validates Theorems 2 and 4. It represents one of the crucial driving forces behind the restriction of initial service demand on service platforms like the Huawei Cloud Computing Platform and the Xinhua Internet Hospital. Interestingly, when the service platform increases the restrictions on initial demand, the service price gradually converges as the operation continues, while the gap in expected revenue under the two pricing strategies remains stable within a certain range. It indicates that the long-term effect of initial demand restriction on the service price gradually weakens, while the long-term effect on the expected revenue is relatively stable.



(**a**) (**b**)

Figure 7. Service price comparison between the two service pricing strategies: (**a**) the impact of the initial demand constraint on the optimal service price; (**b**) the impact of the initial demand constraint on the expected revenue.

In conclusion, some core findings still hold in the extension section, which further demonstrate the robustness of the research conclusion.

8. Conclusions

Service platforms represent a critical pathway for the transformation and upgrading of service industries in the digital economy. The simultaneous involvement of full- and part-time service providers, the time-varying service demands of customers, and the initial demand constraints exacerbate the complexity of the operational management of service platforms. In reality, service platforms commonly employ unified pricing or differential pricing strategies to directly regulate the platform's status and thereby achieve operational goals. However, the majority of existing research on service platform pricing has been conducted from a static equilibrium perspective and overlooks the profound influence exerted by the time-varying nature of service demand on optimal pricing.

By distilling the common features inherent in the actual operations of service platforms across various industries, this study constructs a general dynamic pricing model for a revenue-maximizing service platform using optimal control theory. Specifically, this study determines and analyzes the optimal pricing scheme and expected revenues under scenarios with and without initial demand constraints, as well as for the two distinct pricing strategies. Moreover, it conducts both theoretical and numerical comparisons of the disparities in optimal pricing schemes and expected revenues between these two pricing strategies. Finally, by incorporating service quality as an endogenous variable, the fundamental dynamic pricing model is extended. This extension serves to further validate the robustness of some key research findings. Overall, this study makes significant theoretical contributions and holds substantial practical value for the operational management of service platforms.

8.1. Theoretical Contributions

This study can enrich the theoretical knowledge in the field of scientific pricing of service platforms and make several important contributions to the literature on dynamic pricing of service systems, unified pricing vs. differential pricing, and the pricing for a gig economy service platform.

Regarding the dynamic pricing of service systems, this study extends the single service situation without initial demand constraints to a two-service competition situation with initial demand constraints, and solves for and analyzes the optimal dynamic pricing. This study reveals that initial demand constraints exert a negative impact on the optimal service price and a positive impact on the expected revenues, yet their long-term effects are both limited. When the service quality elasticity coefficient exceeds a certain threshold, the service quality has a positive and a negative effect on the optimal unified price and the optimal differential price, respectively. Blind investment in service quality does not necessarily increase service price and expected revenue.

Regarding the unified pricing vs. differential pricing in service systems, this study extends the static pricing scenario of tangible goods to the dynamic pricing scenario of intangible services. It compares the differences in optimal service prices and expected revenues of the service platform under the two pricing strategies, and presents the applicable conditions for the two pricing strategies. This study determines that, during periods of surging service demand, when the revenue-sharing ratio is smaller or the difference in service quality is greater, the service platform should adopt the differential pricing strategy to achieve higher operational revenue. Otherwise, it should adopt the unified pricing strategy. Meanwhile, an over-sensitivity of customers to service prices proves detrimental to promoting the expected revenue.

Regarding the pricing for a gig economy service platform, this study formulates a dynamic pricing model for the service platform based on optimal control theory. To a certain extent, it transcends the limitations inherent in the equilibrium pricing model. This

study shows that the optimal dynamic price rises as the potential customers' demand expands. When the service price elasticity coefficient falls below a certain threshold, the gap between the optimal service prices under unified and differentiated pricing strategies narrows as potential demand and the demand fluctuation factor increase, while the gap in expected revenues between the two pricing strategies widens with the intensification of price competition.

8.2. Practical Implications

The research conclusions can serve platform managers and policymakers with actionable insights as follows.

The conclusions revealed by this study can assist service platform managers in understanding how initial demand restriction and service competition affect optimal dynamic pricing and expected revenue, which provides support for service pricing decision-making. Under the circumstances of initial demand constraints and two service types in competition, there exists an optimal pricing scheme for service platforms. In actual operations, service platform managers should exercise caution when implementing the "hunger marketing" strategy, that is, restricting initial demand. Specifically, to increase service prices, managers should impose stringent limitations on initial demand. To achieve higher operational revenue, managers are advised to relax the initial demand constraints. Service platform managers can also adjust the optimal service price by modifying service quality, and the specific adjustment approach depends on the pricing strategy adopted. When investing in service quality improvements, managers must conduct rigorous cost–benefit analyses to ensure that the gains from enhanced customer satisfaction outweigh the associated increases in operational expenditure.

The conclusions derived from this study can provide a theoretical foundation for service platform managers to select appropriate pricing strategies. Service platform managers should fully leverage the advantages of AI technology to accurately capture customer market trends and their own operational status, and then choose suitable pricing strategies to effectively achieve operational revenue goals. To capitalize on surging demand, for example, medical service platforms that simultaneously operate multiple types of services with significant differences in service quality should adopt a differential pricing strategy. Similarly, takeout service platforms, which usually set a relatively low revenue-sharing ratio, should also adopt a differential pricing strategy. In contrast, ride-hailing platforms, which typically offer similar ride-hailing services with a relatively high revenue-sharing ratio, may be better off adopting a unified pricing strategy. Meanwhile, service platforms can exert influence on and even mitigate customers' sensitivity to service prices through measures such as targeted marketing (issuing coupons), improving after-sales service, and furnishing complimentary additional services.

The conclusions drawn from this study can offer theoretical insights for policymakers to regulate the service market. Policymakers are empowered to implement macroscopic regulatory measures on the service market via scientific and efficient policy instruments. Specifically, these efforts can be directed towards cultivating potential customers, offering policy-based support, steering the trends of the customer market, and standardizing service competition. By leveraging advanced data-mining tools, policymakers can closely monitor customers' sensitivity to service prices and comprehensively analyze the developmental status of the service market. Subsequently, through modulating the potential service demand and the intensity of service competition, they can mitigate the disparities in service prices and operational revenues that arise from diverse pricing strategies adopted by service platforms. It, in turn, contributes to fostering the sound development of the service market.

8.3. Limitations and Future Directions

This study has certain limitations in theoretical modelling and numerical analysis. The formulation of the service platform dynamic pricing model in this study heavily relies on fundamental assumptions. Some of these assumptions overly simplify the real-world operational scenarios of service platforms. In the theoretical modelling, this study assumes that all customers are homogeneous and perceive no difference in the effectiveness of services provided by full-time and part-time service providers, that service providers can respond to customers in a timely manner, and the revenue-sharing ratio remains constant throughout the entire operational cycle. In the numerical analysis, the determination of key parameter values does not comprehensively draw on the real-world operational data of service platforms. Consequently, there is a lack of empirical validation for the crucial findings and conclusions of this research.

Future research could incorporate the heterogeneity in the customer market, the stochastic nature of supply and demand, and the endogeneity of the revenue-sharing ratio to further enhance the theoretical model developed in this study. The theoretical model established herein could also be further extended to accommodate complex real-world situations, such as continuous service quality management, the operation of more than two service types, and competition among service platforms. Moreover, real-world platform data could be collected to verify the accuracy of the constructed model's predictions and the robustness of the research conclusions.

Author Contributions: Conceptualization, J.L. and J.Z.; methodology, J.L.; software, J.S.; validation, J.L., J.S. and J.Z.; formal analysis, J.L.; investigation, J.L. and J.S.; resources, J.Z.; data curation, J.L. and J.S.; writing—original draft preparation, J.L.; writing—review and editing, J.L. and J.Z.; visualization, J.S.; supervision, J.Z.; project administration, J.Z.; funding acquisition, J.L. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by grants from the National Natural Science Foundation of China (No. 71971156, No. 72371188, and No. 72071063), the Youth Fund Project of the Social Sciences Research of the Ministry of Education (No. 24YJC910011), the Natural Science Foundation of Anhui Province (No. 2408085J041), and the Foundation of Shanghai Polytechnic University (No. EGD25QD29).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

Acknowledgments: The authors sincerely thank the editors and the anonymous reviewers for their valuable and constructive suggestions for improving this paper. The authors especially acknowledge Dongxiao Gu, affiliated with the School of Management, Hefei University of Technology.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1

Proof of Lemma 1. The second-order partial derivative of H^{DU} with the service price is as follows:

$$\frac{\partial^2 H^{DU}}{\partial p^2(t)} = 2[\alpha_1(\theta_1 - 1) + \alpha_2(\theta_2 - 1)] < 0$$

□

Appendix A.2

Proof of Proposition 1. According to Pontryagin's extreme value principle and Lemma 1, the optimal dynamic service price $p^*(t)$, optimal demand trajectory $D^*(t)$, and Lagrange multiplier λ have the following relationships:

$$D^*(t) = \frac{\partial H^{DU}}{\partial \lambda} \tag{A1}$$

$$\frac{\partial H^{DU}}{\partial p^*(t)} = 0 \tag{A2}$$

$$\dot{\lambda}^* = \frac{\partial H^{DU}}{\partial D} \tag{A3}$$

Based on Equations (A1) and (A2), the optimal dynamic service price can be calculated as follows:

$$p^*(t) = \frac{K + ae^{bt}(k_1 + k_2) + \lambda(\alpha_1 + \alpha_2)}{2(\alpha_1 k_1 + \alpha_2 k_2)} \tag{A4}$$

With Formula (A3), the Lagrange multiplier can be determined, and substituted to obtain the optimal unified price. Further substituting the price into Equation (8), the total profit of the service platform under a unified pricing strategy can be obtained. □

Appendix A.3

Proof of Theorem 1. When Condition 1 holds, the calculation is as follows:

$$\frac{\partial p^{DU}(t)}{\partial t} = \frac{e^{-t}[abe^{t+bt}(k_1 + k_2) - C(\alpha_1 + \alpha_2)]}{\alpha_1 k_1 - \alpha_2(k - k_2)} > 0$$

□

Appendix A.4

Proof of Corollary 1. When $0 < \theta_2 < \theta_1$, $\frac{\partial p^{DU}(t)}{\partial \beta_1} = \frac{k_1 q_1}{\alpha_1 k_1 + \alpha_2(k-k_2)} > 0$, $\frac{\partial p^{DU}(t)}{\partial \beta_2} = \frac{k_1 q_2}{\alpha_1 k_1 + \alpha_2(k-k_2)} > 0$, $\frac{\partial p^{DU}(t)}{\partial q_2} = \frac{\eta k - \beta_2 k_2}{\alpha_1 k_1 + \alpha_2(k-k_2)} > 0$, $\frac{\partial p^{DU}(t)}{\partial \alpha} = \frac{e^{bt}(k_1+k_2)}{\alpha_1 k_1 - \alpha_2(k-k_2)} > 0$, $\frac{\partial p^{DU}(t)}{\partial b} = \frac{ae^{bt}t(\theta_1+\theta_2-2)}{\alpha_1 k_1 - \alpha_2(k-k_2)} > 0$. $\frac{\partial \pi^{DU}}{\partial \rho_1} = \frac{-Tq_1^2}{2} < 0$, $\frac{\partial \pi^{DU}}{\partial \rho_2} = \frac{-Tq_2^2}{2} < 0$, $\frac{\partial \pi^{DU}}{\partial w} = -T < 0$. Meanwhile, if $0 < \beta_1 < \max\{1, \frac{\eta k}{k_1}\}$, then $\frac{\partial p^{DU}(t)}{\partial q_1} = \frac{\beta_1 k_1 + \eta k}{\alpha_1 k_1 + \alpha_2(k-k_2)} < 0$. □

Appendix A.5

Proof of Proposition 2. Let $t = 0$ and substitute Equation (10) into the state equation $Q(t)$. Under the constraint of consumer service demand f, the constant coefficient C can be determined, and then $p^{DU}(t)$ and π^{DU} can be calculated. □

Appendix A.6

Proof of Theorem 2. When Condition 2 is held, the calculation is as follows:

$$\frac{\partial p^{DU}(t)}{\partial t} = \frac{e^{-t}\{f[\alpha_1 k_1 - \alpha_2(k - k_2)] + A - ak(\alpha_1 - 3\alpha_2) + abB\}}{(\alpha_1 + \alpha_2)[\alpha_1 k_1 - \alpha_2(k - k_2)]} > 0$$

$$\frac{\partial^2 p^{DU}(t)}{\partial t^2} = \frac{e^{-t}\{ak(\alpha_1 - 3\alpha_2) + ab^2 B + A + f[\alpha_1 + \alpha_2(1 - 2\theta_2) + \theta_1(\alpha_2 - \alpha_1)]\}}{(\alpha_1 + \alpha_2)[\alpha_1 k_1 - \alpha_2(k - k_2)]} > 0$$

$$\frac{\partial p^{DU}(t)}{\partial f} = \frac{-e^{-t}}{\alpha_1 + \alpha_2} < 0$$

where $A = k[2\beta_1 q_1 \alpha_2 + \beta_2 q_2(\alpha_2 - \alpha_1) + \eta(q_1 - q_2)(\alpha_2 + \alpha_1)]$, $B = e^{t+bt}(\alpha_1 + \alpha_2)(k_1 + k_2)$. □

Appendix A.7

Proof of Corollary 2.

$$\frac{\partial \pi^{DU}}{\partial T} = \frac{e^{-2T} f^2(\alpha_1 k_1 + \alpha_2 k_2)}{(\alpha_1+\alpha_2)^2} + \frac{e^{-2T}(\alpha_1 k_1 + \alpha_1 k_1)[a(\alpha_1 - 3\alpha_2) + A][a(\alpha_1 - 3\alpha_2) + A - 2f]}{(\alpha_1+\alpha_2)^2[\alpha_1 k_1 - \alpha_2(k - k_2)]} + \frac{ae^{(b-1)T} f k_1(k_1+k_2) - f k_1 K[\sinh(T) - \cosh(T)]}{\alpha_2(k-k_2) - \alpha_1 k_1} + \frac{e^{-T} k_1 K[a(\alpha_1 - 3\alpha_2) + A] + \alpha_2 k[K^2 + a^2 e^{2bT}(k_1+k_2)^2]}{[\alpha_2(k-k_2) - \alpha_1 k_1]^2} + \frac{ae^{(b-1)T}(k_1+k_2)[ak_1(\alpha_1 - 3\alpha_2) + A] + 2e^T \alpha_2 K}{[\alpha_2(k-k_2) - \alpha_1 k_1]^2} - \frac{\rho_1 q_1^2 + \rho_2 q_2^2}{2}$$

Because the coefficient of $\frac{\partial \pi^{DU}}{\partial T}$ about f is greater than 0, it can be obtained $\frac{\partial \pi^{DU}}{\partial T} > 0$ when Condition 3 holds. □

Appendix B

Appendix B.1

Proof of Lemma 2. The Hessian matrix of H^{DD} regarding decision variables is as follows:

$$S = \begin{pmatrix} \frac{\partial^2 H^{DD}}{\partial p_1^2} & \frac{\partial^2 H^{DD}}{\partial p_1 \partial p_2} \\ \frac{\partial^2 H^{DD}}{\partial p_2 \partial p_1} & \frac{\partial^2 H^{DD}}{\partial p_2^2} \end{pmatrix} = \begin{pmatrix} 2(\alpha_1 + \gamma)(\theta_1 - 1) & -\gamma(2 + \theta_1 + \theta_2) \\ -\gamma(2 + \theta_1 + \theta_2) & 2(\alpha_2 + \gamma)(\theta_2 - 1) \end{pmatrix}$$

The principal minor sequences of S satisfy the following,

$$S_1 = 2(\alpha_1 + \gamma)(\theta_1 - 1) < 0$$

$$S_2 = 2(\alpha_1 + \gamma)(\alpha_2 + \gamma)(\theta_1 - 1)(\theta_2 - 1) - \gamma^2(\theta_1 + \theta_2 - 2)^2 > 0$$

which means that S is negative. Thus, there is an optimal decision-making scheme for the service platform. □

Appendix B.2

Proof of Proposition 3. According to Pontryagin's extreme value principle and Lemma 2, we derive the optimal dynamic service price $\boldsymbol{p}^*(t)$ and the Lagrange multiplier λ as follows:

$$p_1(t) = \frac{-A_1 A_2}{B} \tag{A5}$$

$$p_2(t) = \frac{A_3 + A_4 - A_5}{B} \tag{A6}$$

$$\lambda_1 = k_1 p_1(t) + C_1 e^{-t} \tag{A7}$$

$$\lambda_2 = k_2 p_2(t) + C_2 e^{-t} \tag{A8}$$

where $A_1 = 2k_2 R_2[ak_1 e^{bt} + k_1 q_1(\beta_1 + \eta) + \lambda_1 R_1 - \eta k_1 q_2 - \gamma \lambda_2] - \gamma(k_1 + k_2)$

$$A_2 = \eta q_1 k_2 + \gamma \lambda_1 - ak_2 e^{bt} - k_2 q_2(\beta_2 + \eta) - \lambda_2 R_2$$

$$A_3 = k_1 q_1[\gamma \beta_1(k_1 + k_2) + \eta(2\alpha_1 + \gamma\theta_1 - 2\alpha_1\theta_2 - \gamma\theta_2)]$$

$$A_4 = k_1q_2[2\gamma k_2\beta_2 + 2\alpha_1k_2(\beta_2+\eta) - \gamma k\eta] + ak_1e^{bt}[2\alpha_1k_2 + \gamma(k_1+3k_2)]$$
$$A_5 = \gamma\theta_1\lambda_1(\alpha_1-\gamma) + \gamma\lambda_1R_1\theta_2 + \lambda_2[2\alpha_1k_1R_2 + \gamma(2k_1\alpha_2+\gamma k)]$$
$$B = \gamma[\gamma k^2 + 4\alpha_2(k_1+\theta_2-\theta_1\theta_2)] - 4\alpha_1KR_2$$

We calculate the optimal differentiated prices $\boldsymbol{p^{DD}}(t)$ and substitute them into Equation (12) to obtain the expected revenue of the service platform under the differentiated pricing strategy. □

Appendix B.3

Proof of Theorem 3. If Condition 5 holds, then the following can be calculated:

$$\frac{\partial p_1^{DD}(t)}{\partial t} = \frac{e^{-t}[abe^{t+bt}Z(\alpha_2+2\gamma) + C_1(\alpha_1\alpha_2+\gamma\alpha_1+\gamma\alpha_2+\theta_1\gamma^2-\theta_2R_1R_2) - \gamma kC_2R_1]}{Z(\gamma\alpha_2+\alpha_1R_2)} < 0$$

□

Appendix B.4

Proof of Corollary 3. $\frac{\partial p_1^{DD}(t)}{\partial q_1} = \frac{\gamma\beta_1+\alpha_2(\beta_1+\eta)}{\gamma\alpha_2+\alpha_1(\alpha_2+\gamma)} > 0$, $\frac{\partial p_1^{DD}(t)}{\partial \beta_1} = \frac{q_1(\alpha_2+\gamma)}{\gamma\alpha_2+\alpha_1(\alpha_2+\gamma)} > 0$, $\frac{\partial p_1^{DD}(t)}{\partial \beta_2} = \frac{\gamma q_2}{\gamma\alpha_2+\alpha_1(\alpha_2+\gamma)} > 0$, $\frac{\partial p_1^{DD}(t)}{\partial \alpha} = \frac{e^{bt}(\alpha_2+2\gamma)}{\gamma\alpha_2+\alpha_1(\alpha_2+\gamma)} > 0$, $\frac{\partial p_1^{DD}(t)}{\partial b} = \frac{ae^{bt}t(\alpha_2+2\gamma)}{\gamma\alpha_2+\alpha_1(\alpha_2+\gamma)} > 0$, $\frac{\partial \pi^{DD}}{\partial \rho_1} = \frac{-Tq_1^2}{2} < 0$, $\frac{\partial \pi^{DD}}{\partial \rho_2} = \frac{-Tq_2^2}{2} < 0$, $\frac{\partial \pi^{DD}}{\partial w} = -T < 0$. If $0 < \beta_2 < \min\{1, \frac{\eta\alpha_2}{\gamma}\}$, then $\frac{\partial p_1^{DD}(t)}{\partial q_2} = \frac{\gamma\beta_2-\eta\alpha_2}{\gamma\alpha_2+\alpha_1(\alpha_2+\gamma)} > 0$; if $q_1 < q_2$, then $\frac{\partial p_1^{DD}(t)}{\partial \eta} = \frac{\alpha_2(q_1-q_2)}{\gamma\alpha_2+\alpha_1(\alpha_2+\gamma)} > 0$. □

Appendix B.5

Proof of Proposition 4. Let $t = 0$ and substitute Equations (14) and (15) into the state equation $\boldsymbol{Q}(t) = \{Q_1(t), Q_2(t)\}$. Under the constraint of customer initial service demand, the constant coefficient can be determined, $C_1 = \frac{f_1k_1R_2+\gamma f_2k_2}{\gamma\alpha_2+\alpha_1R_2}$, $C_2 = \frac{f_2k_2R_1+\gamma f_1k_1}{\gamma\alpha_2+\alpha_1R_2}$, and then $p^{DD}(t)$ and π^{DD} can be calculated. □

Appendix B.6

Proof of Theorem 4. When service demand has a surge trend, the following can be calculated:

$$\frac{\partial p_1^{DD}(t)}{\partial t} = \frac{e^{-t}[\gamma f_2 + f_1R_2 + abe^{t+bt}(\alpha_2+2\gamma)]}{\gamma\alpha_2+\alpha_1R_2} > 0,\ \frac{\partial p_1^{DD}(t)}{\partial f_1} = \frac{-e^{-t}R_2}{\gamma\alpha_2+\alpha_1R_2} < 0,\ \frac{\partial p_1^{DD}(t)}{\partial f_2} = \frac{-e^{-t}\gamma}{\gamma\alpha_2+\alpha_1R_2} < 0$$

□

Appendix B.7

Proof of Corollary 4. Since the exponential function, e^x is a monotone function, and therefore T^* satisfies, as shown in the following calculation:

$$\frac{\partial \pi^{DD}}{\partial T} = \frac{e^{-2T^*}}{2(\gamma\alpha_2+\alpha_1R_2)}\{-2ae^{T^*+bT^*}f_1k_1(\alpha_2+2\gamma) + f_2k_2(\alpha_1+2\gamma) - 2k_1e^{T^*}[f_1(\alpha_2q_1\beta_1+\gamma q_1\beta_1+\gamma q_2\beta_2+\alpha_2\eta(q_1-q_2)) + f_2k_2(\gamma q_1\beta_1+q_2\beta_2R_1+\alpha_1\eta(q_2-q_1))] + 2[k_1R_2f_1^2 + k_2R_1f_2^2 + \gamma f_1f_2(k_1+k_2)] + e^{2T^*}(\alpha_2\gamma+\alpha_1R_2)(2w+\rho_1q_1^2+\rho_2q_2^2)\} = 0$$

If Condition 6 holds, then the calculation is as follows:

$$\frac{\partial^2 \pi^{DD}}{\partial T^2} = \frac{e^{-2T}}{\gamma\alpha_2+\alpha_1 R_2}\{k_1 f_1[e^T(\beta_1 q_1\alpha_2 + \gamma\beta_1 q_1 + \gamma\beta_2 q_2 + \eta\alpha_2(q_1-q_2)) - 2f_1 R_2] + ae^{T+bT}(1-b)[k_1 f_1(\alpha_2+2\gamma) + k_2 f_2(\alpha_1+2\gamma)] - 2k_2 R_1 f_2^2 + k_2 f_2 e^T[\gamma\beta_1 q_1 + \beta_2 q_2 R_1 + \alpha_1\eta(q_2-q_1)] - 2\gamma f_1 f_2(k_1+k_2)\} < 0$$

Thus, π^{DD} is a concave function with T, and T^* is the maximum point of π^{DD}. □

Appendix C

Appendix C.1

Proof of Corollary 5.

$$p^{DU}(t) - p_1^{DD}(t) = \frac{-e^{-t}}{(\gamma\alpha_2+\alpha_1 R_2)Z}\{\gamma k R_2 C_2 - C_1[\alpha_1 R_2 + \gamma(\alpha_2+\lambda\theta_1)] + ae^{t+bt}(\alpha_2+2\gamma)Z + e^t Z(\beta_1 q_1 R_2 + \eta q_1\alpha_2 + q_2(\beta_2\gamma - \alpha_2\eta)) + \theta_2 R_1 R_2 C_1] + \frac{C(\alpha_1+\alpha_2)}{\alpha_1 k_1 - \alpha_2(k-k_2)} + \frac{Ke^t + ae^{t+bt}(k_1+k_2)}{\alpha_1 k_1 - \alpha_2(k-k_2)}\}$$

If Condition 7 holds, then the following can be calculated:

$$\frac{\partial(p^{DU}(t) - p_1^{DD}(t))}{\partial a} = e^{bt}[\frac{k_1+k_2}{\alpha_1 k_1 - \alpha_2(k-k_2)} - \frac{\alpha_2+2\gamma}{\gamma\alpha_2+\alpha_1 R_2}] < 0$$

$$\frac{\partial(p^{DU}(t) - p_1^{DD}(t))}{\partial b} = ae^{bt}t[\frac{k_1+k_2}{\alpha_1 k_1 - \alpha_2(k-k_2)} - \frac{\alpha_2+2\gamma}{\gamma\alpha_2+\alpha_1 R_2}] < 0$$

□

Appendix C.2

Proof of Corollary 6.

$$\pi^{DU} - \pi^{DD} = \frac{a(C_1+C_2)(e^{(b-1)T}-1)}{1-b} - \frac{e^{-T}[k_1 R_2 C_2^2 + k_2 R_1 C_1^2 - \gamma C_1 C_2(k_1+k_2)]\sinh(T)}{Z} - [\beta_1 q_1 C_1 + \beta_2 q_2 C_2 + \eta(C_1-C_2)(q_1-q_2)] (1-\cosh(T)+\sinh(T)) + \frac{1}{2(\alpha_1 k_1 - \alpha_2(k-k_2))^2}\{\frac{\alpha_2 k(k_1+k_2)^2 a^2(e^{2bT}-1)}{b} + \frac{2a}{b(b-1)}(k_1+k_2)[-2k\alpha_2(e^{bT}-1) + b(k_1 C(e^{(b-1)T}-1)(\alpha_1+\alpha_2)^2 + 2k\alpha_2(e^{bT}-1))] + 2[e^{-T}C(e^T-1)(\alpha_1+\alpha_2)^2 + \alpha_2 T k^2 + e^{-T}\sinh(T)(\alpha_1 k_1 + \alpha_2 k_2)C^2(\alpha_1+\alpha_2)^2]\}$$

If Condition 8 holds, then the following can be calculated:

$$\frac{\partial(\pi^{DU} - \pi^{DD})}{\partial\gamma} = \frac{e^{-T}(C_1-C_2)(C_1+k_1 C_2 - \theta_2 C_1)\sinh(T)}{Z} > 0$$

□

Appendix D

Appendix D.1

Proof of Proposition 5. The Hessian matrix of H^{DU} regarding decision variables is as follows:

$$S^{DU} = \begin{pmatrix} \frac{\partial^2 H^{DU}}{\partial p^2} & \frac{\partial^2 H^{DU}}{\partial p \partial s_1} & \frac{\partial^2 H^{DU}}{\partial p \partial s_2} \\ \frac{\partial^2 H^{DU}}{\partial s_1 \partial p} & \frac{\partial^2 H^{DU}}{\partial s_1^2} & \frac{\partial^2 H^{DU}}{\partial s_1 \partial s_2} \\ \frac{\partial^2 H^{DU}}{\partial s_2 \partial p} & \frac{\partial^2 H^{DU}}{\partial s_2 \partial s_1} & \frac{\partial^2 H^{DU}}{\partial s_2^2} \end{pmatrix} = \begin{pmatrix} 2\alpha(\theta_1+\theta_2-2) & 0 & 0 \\ 0 & -\rho_1 & 0 \\ 0 & 0 & -\rho_2 \end{pmatrix}$$

The principal minor sequences of S^{DU} satisfy the following,

$$S_1^{DU} = 2\alpha(\theta_1+\theta_2-2) < 0$$

$$S_2^{DU} = 2\alpha(2-\theta_1-\theta_2)\rho_1 > 0$$

$$S_3^{DU} = 2\alpha(\theta_1 + \theta_2 - 2)\rho_1\rho_2 < 0$$

which means that S^{DU} is negative. Thus, there is an optimal decision-making scheme for the service platform. According to Pontryagin's extreme value principle, we derive the optimal dynamic service price $p(t)$ and Lagrange multiplier $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \lambda_3\}$ under a unified pricing strategy. □

Appendix D.2

Proof of Corollary 7.

$$\frac{\partial p(t)}{\partial \alpha} = \frac{[q_1(t)+q_2(t)]\beta+\eta\theta_1 q_2(t)-\theta_1 R_3 q_1(t)+\eta\theta_2 q_2(t)-R_3\theta_2 q_2(t)-ae^{bt}(\theta_1+\theta_2-2)}{2\alpha^2(\theta_1+\theta_2-2)} < 0$$
$$\frac{\partial p(t)}{\partial \beta} = \frac{k_1 q_1(t)+k_2 q_2(t)}{2\alpha(\theta_1+\theta_2-2)} > 0,\ \frac{\partial p(t)}{\partial t} = \frac{abe^{bt}}{2\alpha} > 0,\ \frac{\partial p(t)}{\partial a} = \frac{e^{bt}}{2\alpha} > 0,\ \frac{\partial p(t)}{\partial b} = \frac{ae^{bt}t}{2\alpha} > 0$$

When $\beta > \{\frac{-\eta k}{k_1}, \frac{\eta k}{k_2}\}$, the following can be calculated:

$$\frac{\partial p(t)}{\partial q_1} = \frac{\beta k_1 + \eta k}{2\alpha(\theta_1 + \theta_2 - 2)} > 0, \frac{\partial p(t)}{\partial q_2} = \frac{\beta k_2 - \eta k}{2\alpha(\theta_1 + \theta_2 - 2)} > 0$$

When $q_1(t) < q_2(t)$, the following can be calculated:

$$\frac{\partial p(t)}{\partial \theta_1} = \frac{[q_1(t) - q_2(t)](\beta + 2\eta)k_2}{2\alpha(\theta_1 + \theta_2 - 2)^2} > 0,\ \frac{\partial p(t)}{\partial \theta_2} = \frac{-[q_1(t) - q_2(t)](\beta + 2\eta)k_1}{2\alpha(\theta_1 + \theta_2 - 2)^2} < 0$$

□

Appendix D.3

Proof of Proposition 6. The Hessian matrix of H^{DD} regarding decision variables is as follows:

$$S^{DD} = \begin{pmatrix} \frac{\partial^2 H^{DD}}{\partial p_1^2} & \frac{\partial^2 H^{DD}}{\partial p_1 \partial p_2} & \frac{\partial^2 H^{DD}}{\partial p_1 \partial s_1} & \frac{\partial^2 H^{DD}}{\partial p_1 \partial s_2} \\ \frac{\partial^2 H^{DD}}{\partial p_2 \partial p_1} & \frac{\partial^2 H^{DD}}{\partial p_2^2} & \frac{\partial^2 H^{DD}}{\partial p_2 \partial s_1} & \frac{\partial^2 H^{DD}}{\partial p_2 \partial s_2} \\ \frac{\partial^2 H^{DD}}{\partial s_1 \partial p_1} & \frac{\partial^2 H^{DD}}{\partial s_1 \partial p_2} & \frac{\partial^2 H^{DD}}{\partial s_1^2} & \frac{\partial^2 H^{DD}}{\partial s_1 \partial s_2} \\ \frac{\partial^2 H^{DD}}{\partial s_2 \partial p_1} & \frac{\partial^2 H^{DD}}{\partial s_2 \partial p_2} & \frac{\partial^2 H^{DD}}{\partial s_2 \partial s_1} & \frac{\partial^2 H^{DD}}{\partial s_2^2} \end{pmatrix}$$
$$= \begin{pmatrix} 2k_1 R & \gamma(2 - \theta_1 - \theta_2) & 0 & 0 \\ \gamma(2 - \theta_1 - \theta_2) & 2k_2 R & 0 & 0 \\ 0 & 0 & -\rho_1 & 0 \\ 0 & 0 & 0 & -\rho_2 \end{pmatrix}$$

The principal minor sequences of S^{DD} hold as follows,

$$S_1^{DD} = 2k_1 R < 0$$

$$S_2^{DD} = 4z\alpha^2 + 8\alpha\gamma z - \gamma^2 k^2 > 0$$

$$S_3^{DD} = \left[\gamma^2(\theta_1 + \theta_2 - 2)^2 - 4R^2 z\right]\rho_1 < 0$$

$$S_4^{DD} = \left(4R^2 z - \gamma^2(\theta_1 + \theta_2 - 2)^2\right)\rho_1\rho_2 > 0$$

which means that S^{DD} is negative. Thus, there is an optimal decision-making scheme for the service platform under the differentiated pricing strategy. According to Pontryagin's

extreme value principle, we derive the optimal dynamic service price $\boldsymbol{p}(t) = \{p_1(t), p_2(t)\}$ and Lagrange multiplier $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ under a differentiated pricing strategy. □

Appendix D.4

Proof of Corollary 8.

$$\frac{\partial p_1(t)}{\partial t} = \frac{abe^{bt}k_2[2\alpha k_1 + \gamma(3\theta_1 + \theta_2 - 4)]}{4z\alpha^2 + 8\alpha\gamma z - \gamma^2 k^2} > 0, \ \frac{\partial p_1(t)}{\partial a} = \frac{e^{bt}k_2[2\alpha k_1 + \gamma(3\theta_1 + \theta_2 - 4)]}{4z\alpha^2 + 8\alpha\gamma z - \gamma^2 k^2} > 0$$

$$\frac{\partial p_1(t)}{\partial b} = \frac{tae^{bt}k_2[2\alpha k_1 + \gamma(3\theta_1 + \theta_2 - 4)]}{4z\alpha^2 + 8\alpha\gamma z - \gamma^2 k^2} > 0$$

When $\alpha > \max\left\{\frac{\gamma(-2\beta k_1 - \eta k)}{2k_1 R_1}, \frac{\gamma[\beta(\theta_1+\theta_2-2)-\eta k]}{2\eta k_1}\right\}$, the following can be calculated:

$$\frac{\partial p_1(t)}{\partial q_1} = \frac{(2\beta\gamma k_1 + 2\alpha k_1 R_3 + \gamma\eta k)k_2}{4z\alpha^2 + 8\alpha\gamma z - \gamma^2 k^2} > 0, \ \frac{\partial p_1(t)}{\partial q_2} = \frac{k_2[2\alpha\eta - (2\alpha + \gamma)\eta\theta_1 + \gamma\eta\theta_2 + \beta\gamma(\theta_1 + \theta_2 - 2)]}{4z\alpha^2 + 8\alpha\gamma z - \gamma^2 k^2} < 0$$

When $\frac{\gamma^2 k^2 + 4\alpha^2(\theta_1+\theta_2-\theta_1\theta_2-1)}{8\alpha\gamma} < z$, the following can be calculated:

$$\frac{\partial p_1(t)}{\partial \beta} = \frac{k_2[2k_1 R_3 q_1 + \gamma q_2(\theta_1 + \theta_2 - 2)]}{8\alpha\gamma z - \gamma^2 k^2 - 4\alpha^2(\theta_1 + \theta_2 - \theta_1\theta_2 - 1)} > 0$$

□

References

1. Bai, J.; So, K.C.; Tang, C.S.; Chen, X.; Wang, H. Coordinating supply and demand on an on-demand service platform with impatient customers. *Manuf. Serv. Oper. Manag.* **2019**, *21*, 556–570. [CrossRef]
2. Gu, D.; Wang, Q.; Chai, Y.; Yang, X.; Zhao, W.; Li, M.; Zolotarev, O.; Xu, Z.; Zhang, G. Identifying the risk factors of allergic rhinitis based on Zhihu comment data using a topic-enhanced word-embedding model: Mixed method study and cluster analysis. *J. Med. Internet Res.* **2024**, *26*, e48324. [CrossRef]
3. Gu, D.; Liu, H.; Zhao, H.; Yang, X.; Li, M.; Liang, C. A deep learning and clustering-based topic consistency modeling framework for matching health information supply and demand. *J. Assoc. Inf. Sci. Technol.* **2024**, *75*, 152–166. [CrossRef]
4. Wu, H.; Lu, N. Service provision, pricing, and patient satisfaction in online health communities. *Int. J. Med. Inform.* **2018**, *110*, 77–89. [CrossRef]
5. Liu, W.; Yan, X.; Wei, W.; Xie, D. Pricing decisions for service platform with provider's threshold participating quantity, value-added service and matching ability. *Transp. Res. Part E Logist. Transp. Rev.* **2019**, *122*, 410–432. [CrossRef]
6. Maglaras, C.; Yao, J.; Zeevi, A. Optimal price and delay differentiation in large-scale queueing systems. *Manag. Sci.* **2018**, *64*, 2427–2444. [CrossRef]
7. Wang, J.P.; Huang, H.J. Operations on an on-demand ride service system with express and limousine. *Transp. Res. Part B Methodol.* **2022**, *155*, 348–373. [CrossRef]
8. Zhong, Y.; Lan, Y.; Chen, Z.; Yang, J. On-demand ride-hailing platforms with heterogeneous quality-sensitive customers: Dedicated system or pooling system? *Transp. Res. Part B Methodol.* **2023**, *173*, 247–266. [CrossRef]
9. Li, J.; Chu, C.; Zhang, J.; Sun, J. Bilateral pricing for Internet healthcare platform: Separating or pooling doctors? *Expert Syst. Appl.* **2025**, *294*, 128717. [CrossRef]
10. Nourinejad, M.; Ramezani, M. Ride-Sourcing modeling and pricing in non-equilibrium two-sided markets. *Transp. Res. Part B Methodol.* **2020**, *132*, 340–357. [CrossRef]
11. Duan, Y.; Cao, Y. Optimal pricing and deteriorating inventory control when inventory stimulates stochastic demand with reference price effect. *RAIRO-Oper. Res.* **2021**, *55*, S1803–S1821. [CrossRef]
12. Zhang, Z.; Zhang, N.; Wang, J. The influencing factors on impulse buying behavior of consumers under the mode of hunger marketing in live commerce. *Sustainability* **2022**, *14*, 2122. [CrossRef]
13. Peng, Z.; Pan, H.; Yuan, R.; Wang, Y. A comparative analysis of risk factors influencing crash severity between full-time and part-time riding-hailing drivers in China. *Transp. Lett.* **2025**, *17*, 612–627. [CrossRef]

14. Liu, T.; Xu, Z.; Vignon, D.; Yin, Y.; Qin, Z.; Li, Q. Threshold-based incentives for ride-sourcing drivers: Implications on supply management and welfare effects. *Transp. Res. Part C Emerg. Technol.* **2023**, *156*, 104323. [CrossRef]
15. Bergman, A.; Song, H.; David, G.; Spetz, J.; Candon, M. The role of schedule volatility in home health nursing turnover. *Med. Care Res. Rev.* **2022**, *79*, 382–393. [CrossRef]
16. Anderson, M.; Lee, J.H.; Mashruwala, R.; Yu, Z. The effects of sales changes on resource utilisation in the airline industry. *Account. Financ.* **2025**. [CrossRef]
17. Bell, C.N.; Owens-Young, J.L.; Thorpe, R.J., Jr. Self-employment, working hours, and hypertension by race/ethnicity in the USA. *J. Racial Ethn. Health Disparities* **2023**, *10*, 2207–2217. [CrossRef]
18. Yonezawa, K.; Gómez, M.I.; Mclaughlin, E.W. Impacts of minimum wage increases in the U.S. retail sector: Full-Time versus part-time employment. *J. Agric. Resour. Econ.* **2022**, *47*, 324–341.
19. Gupta, M.; Behl, A.; Pereira, V.; Yahiaoui, D.; Varma, A. "From full-time to part-time": Motivation model for the turbulence-hit knowledge workers. *J. Bus. Res.* **2023**, *163*, 113926. [CrossRef]
20. Zimmert, F.; Zimmert, M. Part-time subsidies and maternal reemployment: Evidence from a difference-in-differences analysis. *J. Appl. Econom.* **2024**, *39*, 1149–1171. [CrossRef]
21. Bosco, M.G.; Valeriani, E. Part time jobs, fragmentation and work instability: Light on the gender gap in Emilia-Romagna. *Int. J. Manpow.* **2024**, *45*, 614–638. [CrossRef]
22. Beirigo, B.A.; Schulte, F.; Negenborn, R.R. A learning-based optimization approach for autonomous ridesharing platforms with service-level contracts and on-demand hiring of idle vehicles. *Transp. Sci.* **2022**, *56*, 677–703. [CrossRef]
23. Bernstein, F.; DeCroix, G.A.; Keskin, N.B. Competition between two-sided platforms under demand and supply congestion effects. *Manuf. Serv. Oper. Manag.* **2021**, *23*, 1043–1061. [CrossRef]
24. Cachon, G.P.; Daniels, K.M.; Lobel, R. The role of surge pricing on a service platform with self-scheduling capacity. *Manuf. Serv. Oper. Manag.* **2017**, *19*, 368–384. [CrossRef]
25. Lin, X.; Zhou, Y.W. Pricing policy selection for a platform providing vertically differentiated services with self-scheduling capacity. *J. Oper. Res. Soc.* **2019**, *70*, 1203–1218. [CrossRef]
26. Chen, M.; Zhao, D.; Gong, Y.; Rekik, Y. An on-demand service platform with self-scheduling capacity: Uniform versus multiplier-based pricing. *Int. J. Prod. Econ.* **2022**, *243*, 108329. [CrossRef]
27. Hu, B.; Hu, M.; Zhu, H. Surge pricing and two-sided temporal responses in ride hailing. *Manuf. Serv. Oper. Manag.* **2022**, *24*, 91–109. [CrossRef]
28. Chen, X.; Zheng, H.; Ke, J.; Yang, H. Dynamic optimization strategies for on-demand ride services platform: Surge pricing, commission rate, and incentives. *Transp. Res. Part B Methodol.* **2020**, *138*, 23–45. [CrossRef]
29. Mejjaouli, S.; Tadj, L. Optimal pricing in ride-share platforms. *Cogent Engineering.* **2023**, *10*, 2230710. [CrossRef]
30. Yu, S.; Yang, Z.; Zhang, W. Differential pricing strategies of air freight transport carriers in the spot market. *J. Air Transp. Manag.* **2019**, *75*, 9–15. [CrossRef]
31. Miao, R.; Li, Q.; Huang, W.; Guo, P.; Mi, L.; Zhang, Z.; Zhang, J.; Jiang, Z. Profit optimization for mileage-based pricing of electric vehicle lease. *IEEE Trans. Eng. Manag.* **2020**, *69*, 951–962. [CrossRef]
32. Ma, S.; Guo, W. Dynamic pricing in a two-class queueing system with arrival and service rate control. *IMA J. Manag. Math.* **2021**, *32*, 361–380. [CrossRef]
33. Zhao, D.; Yuan, Z.; Chen, M.; Yang, S. Differential pricing strategies of ride-sharing platforms: Choosing customers or drivers? *Int. Trans. Oper. Res.* **2022**, *29*, 1089–1131. [CrossRef]
34. Zeng, S.; Xu, X.; Gu, Y.; Zhang, J. Uniform pricing effect: The influence of uniform pricing on product value judgement and purchase intention. *Int. J. Consum. Stud.* **2023**, *47*, 1364–1376. [CrossRef]
35. Radhi, M.; Zhang, G. Pricing policies for a dual-channel retailer with cross-channel returns. *Comput. Ind. Eng.* **2018**, *119*, 63–75. [CrossRef]
36. Fan, Y.; Yan, X.; Cui, L.; Zhang, L.; Wang, J. Carbon pricing, carbon equity, and the RCEP framework. *China Econ. Rev.* **2023**, *80*, 102017. [CrossRef]
37. Wang, L.; Zhou, L.; Chan, H.K.; Fan, T. Impact of cost structure and cannibalisation on live streaming pricing strategies. *Int. J. Prod. Econ.* **2025**, *285*, 109619. [CrossRef]
38. Gao, J.; Xiao, Z.; Wei, H. Competition and coordination in a dual-channel green supply chain with an eco-label policy. *Comput. Ind. Eng.* **2021**, *153*, 107057. [CrossRef]
39. Ha, A.Y.; Tong, S.; Wang, Y. Channel structures of online retail platforms. *Manuf. Serv. Oper. Manag.* **2022**, *24*, 1547–1561. [CrossRef]
40. Guo, J.; Zhou, Y.; Li, B. The optimal pricing and service strategies of a dual-channel retailer under free riding. *J. Ind. Manag. Optim.* **2022**, *18*, 2049–2076. [CrossRef]
41. Li, G.; Li, L.; Sun, J. Pricing and service effort strategy in a dual-channel supply chain with showrooming effect. *Transp. Res. Part E Logist. Transp. Rev.* **2019**, *126*, 32–48. [CrossRef]

42. Nie, F.; Li, X.; Zhou, C. Impact of privacy regulation involving information collection on the ride-hailing market. *Omega* **2024**, *126*, 103077. [CrossRef]
43. Mo, D.; Wang, H.; Cai, Z.; Szeto, W.Y.; Chen, X.M. Modeling and regulating a ride-sourcing market integrated with vehicle rental services. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *192*, 103797. [CrossRef]
44. Chen, Z.; Miao, Y.; Ke, J.; He, Q.C. Operations and regulations for a ride-sourcing market with a mixed fleet of human drivers and autonomous vehicles. *Transp. Res. Part C Emerg. Technol.* **2024**, *160*, 104519. [CrossRef]
45. Lin, X.; Jin, K.; Lin, Q.; Zhou, Y.W.; Fu, W. Asset-light or asset-heavy? Implications for ride-hailing platforms' profits, consumer surplus, and driver surplus. *Comput. Ind. Eng.* **2024**, *194*, 110336. [CrossRef]
46. Chen, D.; Zhou, Y.W.; Lin, X.; Jin, K. Customer and provider bounded rationality in on-demand service platforms. *Eur. J. Oper. Res.* **2025**, *320*, 389–401. [CrossRef]
47. Gui, Y.; Zhai, H.; Dong, F.; Liu, Z. Value-added service investment and pricing decisions in two-sided platform: The role of user expectations. *Kybernetes* **2023**, *53*, 557–579. [CrossRef]
48. Sui, R.; Zhang, X.; Dan, B.; Zhang, H.; Liu, Y. Bilateral value-added service investment in platform competition with cross-side network effects under multihoming. *Eur. J. Oper. Res.* **2023**, *304*, 952–963. [CrossRef]
49. Zhu, W.; Xie, J.; Xia, Y.; Wei, L.; Liang, L. Getting more third-party participants on board: Optimal pricing and investment decisions in competitive platform ecosystems. *Eur. J. Oper. Res.* **2023**, *307*, 177–192. [CrossRef]
50. Zhong, Y.; Pan, Q.; Xie, W.; Cheng, T.C.E.; Lin, X. Pricing and wage strategies for an on-demand service platform with heterogeneous congestion-sensitive customers. *Int. J. Prod. Econ.* **2020**, *230*, 107901. [CrossRef]
51. Tang, Y.; Guo, P.; Tang, C.S.; Wang, Y. Gender-related operational issues arising from on-demand ride-hailing platforms: Safety concerns and system configuration. *Prod. Oper. Manag.* **2021**, *30*, 3481–3496. [CrossRef]
52. Lei, M.; Liu, S.; Jasin, S.; Vakhutinsky, A. Joint inventory and pricing for a one-warehouse multistore problem: Spiraling phenomena, near optimal policies, and the value of dynamic pricing. *Oper. Res.* **2024**, *72*, 738–762. [CrossRef]
53. Siddiq, A.; Taylor, T.A. Ride-hailing platforms: Competition and autonomous vehicles. *Manuf. Serv. Oper. Manag.* **2022**, *24*, 1511–1528. [CrossRef]
54. Benjaafar, S.; Ding, J.Y.; Kong, G.; Taylor, T. Labor welfare in on-demand service platforms. *Manuf. Serv. Oper. Manag.* **2022**, *24*, 110–124. [CrossRef]
55. Schlosser, R.; Chenavaz, R.Y. Joint dynamic pricing and marketing-mix strategies for revenue management applications with stochastic demand. *Int. Trans. Oper. Res.* **2025**, *32*, 1566–1592. [CrossRef]
56. Jia, Z.; Chen, Q.; Xu, Q. The art of balancing price and plug: Developing a theoretical model for dynamic pricing in the electric vehicle market. *Sustainability* **2024**, *16*, 9325. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

How Doctors' Proactive Crafting Behaviors Influence Performance Outcomes: Evidence from an Online Healthcare Platform

Wenlong Liu *, Yashuo Yuan, Zifan Bai and Shenghui Sang

College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; sx2309082@nuaa.edu.cn (Y.Y.); bzf0506@nuaa.edu.cn (Z.B.); sang.sh@nuaa.edu.cn (S.S.)
* Correspondence: willenliu@nuaa.edu.cn

Abstract

With the steady global progress in integrating technology into healthcare delivery, doctors' behavioral patterns on online healthcare platforms have increasingly become a focal point in the fields of digital health and healthcare service management. Grounded in Job Crafting Theory, this study constructs a proactive crafting index, which captures doctors' proactive behaviors on the platform across three dimensions: consultation rate, number of consultations, and response speed. We systematically examine the multidimensional impacts of such behaviors on performance outcomes, including online consultation volume, offline service volume, and user evaluation performance. This study collects publicly available records from a major online healthcare platform in China and conducts empirical analysis using the entropy weight method and econometric techniques. The results reveal that there is an optimal level of proactive engagement: moderate proactivity maximizes online consultation volume, while both insufficient and excessive proactivity reduce it. Offline service volume, in contrast, follows a U-shaped relationship, where moderate proactive engagement minimizes offline visits, while too little or too much engagement leads to more offline service needs. These nonlinear patterns highlight the importance of framing doctors' proactive behavior to optimize both online engagement and offline service. The findings enrich Job Crafting Theory by identifying boundaries in platform-based service environments and provide actionable insights for platform operators to design behavior management and incentive systems tailored to doctors' professional rank, patient condition, and regional context.

Keywords: online healthcare platform; Job Crafting Theory; doctors' proactive behavior; performance outcomes

1. Introduction

With the rapid advancement of digital health initiatives worldwide, online healthcare platforms have increasingly become an essential component of modern healthcare systems across countries. Integrating telemedicine, remote consultations, and digital patient management, these platforms are transforming traditional models of care delivery, enhancing healthcare accessibility, efficiency, and patient engagement on a global scale [1,2]. Through these platforms, doctors can provide patients with services such as text-based consultations, voice inquiries, and video appointments, enabling efficient remote interactions. In this process, doctors not only perform their traditional clinical duties but also take on roles as service providers, communicators, and representatives of their professional image on the

platform. Their behavioral patterns have become increasingly influential in determining service performance outcomes [2,3]. Compared with offline healthcare, the evaluation of doctors' performance on digital platforms is more transparent, dynamic, and multidimensional, encompassing metrics such as user feedback, repeat visits, and referrals [4]. How to stimulate doctors' proactivity and enhance their overall service performance within platform-based contexts has thus become a focal issue in the operation and management of online healthcare platforms [4]. Recent empirical studies have shown that doctors' response speed, consultation frequency, and communication style significantly impact patient satisfaction, repeat visits, and even platform recommendation exposure. For example, Liu et al. find that better interaction performance attracts more chronic patients [5], while Wang et al. demonstrate that physicians' online activities can lead to a higher service quantity in offline channels, whereas the more offline-patients physicians serve, the more articles the physicians will likely share online [6]. These findings provide strong support for the growing importance of behavioral engagement in platform-based healthcare services.

In recent years, Job Crafting Theory has emerged as a prominent theoretical framework in the fields of organizational behavior and human resource management. The theory posits that individuals are not passive recipients of job assignments; instead, they actively modify the content of their tasks, interpersonal relationships, and role perceptions. Through such proactive adjustments, individuals can acquire more job resources, reduce stressors, and achieve an optimal alignment between job demands and available resources [7,8]. This framework provides a novel perspective for understanding how doctors may self-regulate to influence their performance outcomes on online healthcare platforms [9].

Within these platforms, doctors can actively shape their "work presence" by taking on more consultations, responding promptly, and increasing service frequency, thereby attracting patients to choose their services [10]. To comprehensively understand the synergistic effects among doctors' behavioral indicators, this study develops a novel composite measure—Proactive Crafting Index (PCI)—specifically designed for online healthcare platforms. While grounded in Job Crafting Theory, the PCI is conceptually new and operationally tailored to capture doctors' proactive behaviors in digital service environments. It integrates three key behavioral metrics—consultation rate, number of consultations, and response speed—that reflect task and relational crafting dimensions: specifically, consultation rate and number of consultations correspond to task crafting, while response speed reflects relational crafting, synthesized using the entropy weight method [11,12]. This design enables a multidimensional understanding of doctors' platform engagement, which is often oversimplified in existing research. This index is used to assess doctors' proactive behaviors on online healthcare platforms and to analyze their impact on three dimensions of performance: online consultation volume (OCV), offline service volume (OSV), and user evaluation performance (UEP).

In addition, doctors' performance outcomes may also be influenced by individual attributes and external contextual factors. Variables such as doctors' professional level, the area in which they practice, and the disease urgency level of their patients may moderate the relationship between proactive behaviors and performance outcomes [1,10]. Therefore, this study further investigates the heterogeneity of these variables to construct a more comprehensive research framework. By integrating theory with empirical analysis, this study aims to uncover how doctors' proactive crafting behaviors influence their performance outcomes, thereby offering both theoretical insight and practical guidance for doctor management and innovation in healthcare services on digital platforms.

2. Literature Review

2.1. Studies on Job Crafting Theory

Job Crafting theory was first introduced by Wrzesniewski and Dutton, emphasizing that individuals are not passive recipients of job tasks and role expectations [7]. Instead, they proactively adjust task boundaries, interpersonal relationships, and perceptions of work meaning to align with personal needs and motivations, thereby enhancing job satisfaction and performance. Job crafting behaviors are typically categorized into three forms: task crafting, relational crafting, and cognitive crafting. Later, Tims and Bakker extended the theory within the framework of the Job Demands–Resources (JD-R) Model, suggesting that individuals can enhance challenging work resources and reduce hindering demands through crafting behaviors, which in turn improves performance and alleviates burnout [7,13].

Job crafting theory has been widely applied in research on service industries, knowledge workers, and virtual work environments. Studies have shown that proactive job crafting significantly enhances employee engagement and organizational commitment in public service institutions [1,14]. Among teachers, task and relational crafting behaviors have been linked to increased job satisfaction and innovative behaviors [14]. Chen et al. found that cognitive crafting mitigates feelings of isolation and productivity loss among remote workers, especially in digital platform settings [15]. However, these studies are mostly situated in general organizational or educational contexts and rarely engage with highly institutionalized, regulated environments like healthcare. Furthermore, they often rely on subjective reports rather than directly observing behavior in real-time digital interactions.

In the healthcare context, although work is highly standardized and task-driven, doctors still retain a degree of autonomy during service delivery. They may adjust the pace of service, optimize consultation procedures, and strengthen doctor–patient interactions, thus reshaping the boundaries of their work. Existing studies have shown that task and cognitive crafting behaviors among healthcare professionals contribute to improved work adaptation and well-being [16–18]. Furthermore, individual motivations, departmental characteristics, and service pressures influence whether doctors choose to engage in self-regulation. However, existing research on doctors' crafting behaviors is largely confined to offline or traditional hospital settings. In increasingly digitalized work environments where professionals interact with others through structured systems, algorithmic performance evaluation, and publicly visible profiles, new forms of job crafting may emerge. These settings introduce both novel affordances—such as broader reach and flexible interaction tools—and constraints—such as system rules and transparent performance metrics—that existing job crafting frameworks have yet to fully theorize [19,20]. Therefore, job crafting theory offers a suitable foundation for understanding doctors' behaviors on online healthcare platforms. Investigating job crafting in this context can reveal how doctors actively seek resources, enhance influence, and improve performance outcomes within a highly structured yet dynamic digital framework. Such an investigation deepens our understanding of proactive work behaviors in digital healthcare environments and further explores the boundaries of job crafting theory's applicability beyond traditional settings.

2.2. Studies on Doctors' Behavior

With the rapid development of online healthcare platforms, doctors' behavior on these platforms has become a prominent research topic in healthcare management and service innovation [1,21]. As core service providers, doctors' attitudes, behavioral characteristics, and interaction styles significantly affect patient experience and platform service quality [22,23]. Existing research has explored the relationship between doctors' behaviors and patient responses from various perspectives.

In terms of responsiveness, prompt replies to patient inquiries are seen as key to enhancing trust and satisfaction on the platform [24]. Research findings empirically found a significant negative correlation between doctors' average response time and the likelihood of patient revisit, indicating that response speed plays a critical role in user retention [25]. Regarding service frequency and consultation behavior, doctors with higher service frequency are more likely to receive platform recommendations and attract more patients, forming a positive feedback loop of "high exposure–high traffic–high performance" [26,27]. In addition, doctors' enthusiasm and communication quality have been confirmed to significantly impact service evaluations and reputation building [28]. Other researchers have found that the level of detail in doctors' initial consultations is significantly associated with patient's willingness to continue follow-up consultations [29].

While previous studies have examined the impact of individual behavioral traits on performance, several important limitations remain. First, the majority of existing research tends to focus on single behavioral indicators—such as response speed, frequency, or consultation detail—without capturing the interactive, temporal, and strategic nature of doctors' behavioral patterns. This approach overlooks the possibility that doctors may intentionally and proactively orchestrate multiple behaviors to shape their work environment and outcomes over time [30,31]. Second, most studies are descriptive or correlational in nature, lacking a unifying theoretical framework that can explain why and how certain behavioral configurations emerge and influence performance [31,32]. This approach limits our understanding of proactive behavior as a multidimensional construct—an issue this study aims to address through an integrated analytical model. Given doctors' service autonomy on platforms, their series of continuous, interactive, and adaptive behaviors should be conceptualized as an integrated form of Proactive Crafting Behavior, rather than isolated service actions. Therefore, integrating doctors' behaviors into a job crafting theoretical framework holds substantial theoretical and practical significance.

2.3. Studies on Doctors' Performance

As one of the core quality indicators in healthcare service setting, doctors' performance outcomes have traditionally been assessed based on clinical competence, diagnostic accuracy, and patient's physiological treatment outcomes [33,34]. However, in online healthcare contexts, performance is defined more broadly, encompassing immediacy of service delivery, interaction quality, and patient's perceived outcomes [35,36]. Particularly on internet-based platforms, performance metrics often derive from behavioral logs and user feedback, such as consultation volume, patient ratings, and repeat visits, providing a more holistic picture of doctors' service capabilities and attractiveness to users [23,37]. Among these, user ratings are frequently used to evaluate service quality and platform reputation. Research findings that the positive review rate is a critical factor in platform recommendation algorithms, indirectly influencing a doctors' patient flow and income [22]. Research found a significant positive correlation between doctors' service ratings and patients' willingness to pay [38]. In addition to subjective evaluations, objective behavioral indicators such as Online consultation volume, revisit rates, and Offline service volume are widely adopted to assess doctors' service stickiness and long-term impact on the platform [39–41].

Despite these advancements, current research on online doctors' performance suffers from several conceptual and methodological limitations. First, performance is often operationalized using fragmented or surface-level indicators, such as star ratings or consultation counts, which fail to capture the multidimensional and dynamic nature of service provision on digital platforms [42,43]. Second, most studies treat performance as an outcome influenced primarily by external factors (e.g., platform algorithms, patient preferences),

without systematically investigating the proactive role of doctors themselves in shaping these outcomes [44,45]. Third, theoretical guidance remains weak, as few studies draw on established behavioral or organizational theories to explain how doctors navigate institutional constraints, leverage platform affordances, or align their service strategies with personal or professional goals [43–45]. These gaps highlight the need for a more comprehensive, theory-driven approach to understanding platform-based medical performance. Moreover, few studies have simultaneously examined user evaluation performance, online consultation volume, and offline service volume within a unified analytical framework [46,47], explanatory mechanisms for doctors' performance often focus on platform rules and patient characteristics, with limited attention to doctors' own behavioral motivations and adjustment strategies. Therefore, it is necessary to investigate how doctors achieve multidimensional performance outcomes through proactive behavioral crafting.

2.4. Review Summary

Despite increasing scholarly interest in the relationship between doctors' behavior and performance, several gaps remain: (1) Existing research tends to conceptualize doctors' behaviors in fragmented and static ways—focusing on discrete metrics such as response speed, consultation frequency, or communication quality—without considering their proactive, strategic, and adaptive dimensions over time. This limits our understanding of behavior as a dynamic process shaped by doctors' own agency, especially under platform constraints. (2) Performance outcomes are typically operationalized using single-dimensional indicators such as review ratings or consultation counts, overlooking the multidimensional and interrelated nature of digital healthcare performance. As shown in previous sections, objective (e.g., consultation volume) and subjective (e.g., patient ratings) performance dimensions often co-exist, yet are rarely modeled together. (3) Few studies systematically apply theoretical frameworks to explain the mechanisms linking doctor behavior and performance. While many studies adopt a correlational or descriptive approach, they often lack conceptual tools to interpret why certain behaviors emerge and how they function within the digital institutional context of online platforms.

To address these gaps, this study introduces the concept of a proactive crafting index based on Job Crafting Theory, which comprehensively measures doctors' proactive service behaviors on online healthcare platforms. It constructs a three-dimensional model of performance outcomes encompassing online consultation volume, offline service volume, and user evaluation performance, and explores their interrelationships and moderating mechanisms. This approach aims to enrich the theoretical foundation and empirical evidence for research on doctors' behavior in online healthcare platforms.

3. Research Hypotheses

Based on Job Crafting Theory, individuals in dynamic work environments proactively adjust their work approaches to adapt to external demands, enhance access to resources, and ultimately improve performance outcomes [7,48]. In the context of online healthcare platforms, doctors enjoy relatively high service autonomy, allowing them to reshape their work boundaries through consultation selection, response management, and task coordination [8,49]. These behaviors influence not only how doctors are ranked and presented by platform algorithms, but also play a critical role in shaping patients' perceptions, service experiences, and subsequent choices [50–52]. Therefore, this study conceptualizes doctors' proactive behaviors as a Proactive crafting index and proposes the following hypotheses from a theoretical perspective.

3.1. Proactive Crafting and Online Consultation Volume

The number of online consultations reflects a doctors' service engagement and is shaped by both platform recommendation algorithms and patient choice. According to Job Crafting Theory, proactive behaviors enhance a doctors' exposure and reach and trust score on the platform, increasing the likelihood of being recommended [52]. More frequent consultations and quicker responses are logged by the platform as indicators of high-quality, active engagement, leading to more exposure and patient visits [1]. Furthermore, trust between doctor and patient is a key driver of continued use. Proactive service behavior serves as a trust signal that reduces patients' uncertainty and increases their likelihood of returning to the same doctor [53]. Research found that doctors' proactive behavior significantly improves patient retention on digital platforms [54]. According to the perspective of Job Crafting Theory, excessive job crafting may lead to resource depletion and overload—for instance, if doctors take on too many consultations or excessively shorten response times, it may reduce the quality of each consultation and, in turn, undermine patient trust [14]. Empirical studies have also shown that in online services, users exhibit an implicit threshold for "behavioral frequency": when the number of consultations becomes too high, the likelihood that patients perceive the service as perfunctory also increases [55]. Thus, we propose the following hypothesis:

H1: *Doctors' Proactive crafting index shows an inverted U-shaped relationship with their online consultation volume in the subsequent month.*

3.2. Proactive Crafting and Offline Service Volume

Offline service volume captures the conversion of online consultation into in-person visits, reflecting patients' transfer of trust across channels—a key performance metric [56,57]. In the online healthcare setting, proactive crafting enhances patients' perception of the doctors' professional capabilities, service quality, and communication attitude, thereby increasing their likelihood to visit offline [57,58]. This aligns with Job Crafting Theory, which emphasizes resource acquisition through task boundary expansion and improved interpersonal interactions.

However, the relationship may not be strictly linear. Low to moderate levels of proactive behavior may be perceived by patients as driven more by a desire to boost platform traffic than by a genuine concern for diagnostic quality [59,60]. This perception can undermine patients' trust in a doctor's offline professional competence, thereby reducing their willingness to seek offline consultations [61]. This could deter patients from transitioning to offline visits, especially when these involve additional time and financial costs [62]. In contrast, when proactive behavior reaches a high and consistent level—such as sustained consultation activity accompanied by a stable response rhythm—patients are more likely to interpret it as a signal of professional dedication. They may perceive the doctor as having sufficient time, energy, and expertise to manage both online services and offline care effectively, which can enhance trust in the doctor's offline services and increase the likelihood of offline conversion [63,64]. Therefore, we expect a U-shaped relationship:

H2: *Doctors' Proactive crafting index and their offline service volume in the subsequent month exhibit a U-shaped relationship.*

3.3. Proactive Crafting and User Evaluation Performance

User evaluation performance reflects patients' subjective assessment of the doctors' service and serves as a direct representation of the doctors' reputation on the platform [65]. According to Job Crafting Theory, doctors who proactively adjust their service pace and interaction style can improve timeliness and professionalism, thus enhancing interaction

quality. Tims et al. found that job crafting behaviors improve perceived value and satisfaction among customers [34].

On digital platforms, fast response times, efficient consultation scheduling, and frequent service provision reduce patient waiting time and enhance perceived professionalism. Service enthusiasm and visible dedication further strengthen perceived value, leading to higher satisfaction and better reviews. Studies have shown that proactive behaviors in service-oriented organizations help shape overall customer impressions. Unlike online consultation volume and offline service volume, user evaluation performance (UEP) reflects patients' subjective feedback on service quality. In service scenarios, user evaluations are subject to a "floor effect," meaning they rely more on whether basic expectations are met rather than the rationality or frequency of the doctor's behavior [66,67]. Consequently, the relationship between UEP and doctors' proactive behaviors tends to be more linear. Accordingly, we propose:

H3: *Doctors' Proactive crafting index is positively associated with user evaluation performance.*

3.4. Heterogeneity by Professional Rank, Disease Type, and Regional Context

Although proactive crafting generally enhances performance, this relationship may vary across different subgroups. That is, the effectiveness of proactive behaviors may exhibit structural heterogeneity. From the extended perspective of the Job Demands–Resources (JD-R) model, differences in resource endowment and job context significantly shape the alignment between behavior and performance outcomes [12,68]. Therefore, we examine the heterogeneity across three dimensions: doctor rank, disease urgency, and regional development level.

Doctors at different professional ranks enjoy varied levels of authority and credibility [68]. High-rank doctors are more likely to have their proactive behavior interpreted as dedication and responsibility, while lower-rank doctors might be perceived as promotional or commercially driven. Thus, rank moderates the behavior-performance link.

H4a: *The effect of Proactive Crafting on performance differs significantly by doctors' professional rank: it is stronger in doctors with higher professional ranks than those with lower ranks.*

Patients' sensitivity to responsiveness and initiative differs depending on disease urgency. For acute conditions, prompt responses and proactive communication quickly alleviate patient anxiety, driving repeat visits [69,70]. For chronic diseases, long-term consistency matters more, and proactive behavior may yield diminishing marginal returns.

H4b: *The effect of Proactive Crafting on performance differs significantly across disease types: it is stronger in acute conditions than in non-acute conditions.*

Variations in platform penetration, healthcare infrastructure, and digital literacy exist across regions [71]. In more developed areas, patients are more familiar with online healthcare processes and can better perceive proactive behaviors, leading to positive performance feedback. In underdeveloped regions, limited platform engagement and technological barriers may hinder recognition and effectiveness of proactive behavior.

H4c: *The effect of Proactive Crafting on performance differs significantly across regions: it is stronger in more developed regions compared to less developed regions.*

4. Research Design

4.1. Data Source and Processing

This study is based on publicly accessible secondary data collected from a major Chinese online healthcare platform (www.wedoctor.com) (accessed on 7 June 2023). The platform openly provides non-sensitive information about doctors' online service activities, including consultation frequency, response time, pricing, and patient reviews. We collected data through automated scripts consistent with the platform's robots.txt policy. The dataset comprises more than 100,000 panel records covering over 2800 doctors from December 2022 to May 2023. All doctor- and patient-level data were recorded in the form of platform-generated encrypted IDs, without any personal identifiers such as names, photos, or contact information. Therefore, no personally identifiable information was accessed or stored during data collection. The original behavioral data are generated by the platform as part of standard system logging and were not manipulated by the researchers. To ensure data quality, we performed consistency checks, outlier screening, and cross-validation across time periods. The data are considered reliable as they are systematically recorded by the platform based on real-time user behavior. Because the data are publicly available and do not involve private or sensitive content, this study does not require ethics board approval or informed consent, in line with standard practices for research using non-identifiable secondary data.

During the six-month observation period (December 2022 to May 2023), there were no reported major platform updates, systemic changes, or significant health policy shifts that might systematically bias doctors' behaviors or patient demand. Moreover, seasonal effects—such as those associated with public holidays or flu season—were controlled using time-fixed effects in the regression models to mitigate their potential impact.

4.2. Variable Selection and Measurement

Variable selection is grounded in the core logic of Job Crafting Theory, which posits that employees proactively shape their work to acquire resources, tackle challenges, and improve performance. Drawing on the platform's data structure and doctors' work characteristics, we identify key indicators to construct independent, dependent, and control variables, ensuring theoretical consistency and operational feasibility.

(1) Independent Variable

In the context of Online healthcare platforms, doctors enjoy a high degree of service autonomy. Their proactive behaviors reflect a combination of task crafting and relational crafting within the platform environment. According to Job Crafting theory, employees can enhance resource acquisition and improve performance outcomes by actively expanding work boundaries and optimizing interaction processes.

Given the digital characteristics of doctors' work tasks and Job Crafting Theory, this study selects three behavioral indicators—online consultation rate (t), number of consultations (t), and response speed (t)—as the foundational variables for constructing the proactive crafting index (t). The online consultation rate (t) reflects a doctors' willingness to accept appointment requests, indicating openness to task boundaries, capturing doctors' willingness to proactively expand their service scope [72,73]. The number of consultations (t) measures the supply of services provided by the doctor, representing the ability to expand their work scope, representing doctors' proactive efforts to acquire resources by increasing service volume [74,75]. Response speed (t), defined as the average time a doctor takes to reply to patient inquiries, serves as a key indicator of service agility and interactive proactivity, denoting doctors' proactive efforts to optimize doctor–patient interactions [64,74,75]. Together, these three indicators—whether to accept tasks, how much to invest, and how to optimize interactions—comprehensively capture the key dimen-

sions of doctors' proactive behaviors in online healthcare settings, and are theoretically coherent. They are chosen because they are objectively observable, theoretically grounded, and feasible to extract from platform data. Specifically, consultation rate and number of consultations correspond to task crafting, while response speed reflects relational crafting. In contrast, cognitive crafting, which refers to individuals changing how they perceive the meaning or identity of their work, involves internal psychological processes that are difficult to detect from behavioral traces in a digital setting. While important in theory, cognitive crafting cannot be reliably inferred from platform-level data without self-reported or qualitative inputs. Therefore, our PCI focuses on externally observable proactive behaviors that reflect how doctors shape their task boundaries and interaction processes on the platform.

This study employs the entropy weight method to synthesize the Proactive Crafting Index (PCI), which better fits the research needs compared to equal weighting or principal component analysis (PCA). The equal weighting method assumes all indicators contribute equally, whereas actual data show varying levels of dispersion across the three indicators (e.g., response speed exhibits greater individual variation than consultation rate). The entropy method automatically assigns weights based on information entropy, giving more weight to indicators with higher variability and reducing subjective bias. PCA, by contrast, focuses on dimensionality reduction, which may obscure the unique theoretical meaning of each indicator [76]. In comparison, the entropy method preserves the conceptual integrity of all three indicators—each representing task boundaries, job investment, and interaction quality—making it more appropriate for this measurement task. The resulting composite index reflects the overall tendency of doctors to proactively manage their platform work.

The process involves: standardizing the indicators (using Equation (1) for positive indicators and Equation (2) for negative indicators), calculating the weight of the indicator j for the doctor i (Equation (3)), computing the information entropy of the indicator j (Equation (4)), determining the degree of divergence (Equation (5)), and finally calculating the weight of each indicator (Equation (6)) and the composite score (Equation (7)).

$$y_{ij} = \frac{x_{ij} - min(x_j)}{max(x_j) - min(x_j)} \tag{1}$$

$$y_{ij} = \frac{max(x_j) - x_{ij}}{max(x_j) - min(x_j)} \tag{2}$$

$$p_{ij} = \frac{y_{ij}}{\sum_{i=1}^{n} y_{ij}} \tag{3}$$

$$e_j = -k\sum_{i=1}^{n} p_{ij} \cdot ln(p_{ij}) \tag{4}$$

$$d_j = 1 - e_j \tag{5}$$

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j} \tag{6}$$

$$S_i = \sum_{j=1}^{m} w_j \cdot y_{ij} \tag{7}$$

In Formulas (1)–(7) above, x_{ij} refers to the original (raw) value of indicator j for doctor i. For example, consultation rate, consultation number, or response speed. y_{ij} refers to the normalized value of indicator j for doctor i, obtained through min–max normalization. Equation (1) is used for positive indicators (higher is better), and Equation (2) is used for negative indicators (lower is better, e.g., response time). p_{ij} refers to the proportion

of doctor i's normalized score on indicator j relative to the total across all doctors. e_j refers to the information entropy of indicator j, reflecting the dispersion of this indicator across all individuals. A higher entropy means less variability. d_j refers to the degree of divergence of indicator j, calculated as $1 - e_j$. It reflects the contribution of the indicator to distinguishing between individuals. w_j refers to the entropy-based weight assigned to indicator j, computed by normalizing the divergence across all indicators. S_i refers to the final composite score (PCI) for doctor i, calculated as the weighted sum of the normalized indicators. A higher score reflects stronger proactive crafting behavior.

(2) Dependent Variable

Performance outcomes serve as the core dimension for evaluating the effectiveness of doctors' behaviors on the platform. Traditional performance research often focuses on either subjective satisfaction or a single objective output indicator. However, in online environments, a multidimensional performance structure more accurately captures the comprehensive value of doctors' actions.

Drawing on literature from service marketing and digital platform behavior, this study constructs three performance indicators: user evaluation performance (t+1), online consultation volume (t+1), and offline service volume (t+1). User evaluation performance (t+1) reflects patients' subjective perceptions and satisfaction with the quality of service, measured by the proportion of positive reviews in the current month. Online consultation volume (t+1) indicates patients' behavioral choices to engage the doctors' services in the following month, reflecting both service attractiveness and platform recommendation effects. Offline service volume (t+1) assesses whether online interactions effectively translate into offline medical services, serving as a key outcome for measuring trust transfer and cross-channel integration. This performance framework aligns with the logical chain of service quality → usage behavior → trust conversion, allowing for a comprehensive assessment of the impact of doctors' proactive crafting behaviors on performance outcomes.

(3) Moderating Variables

This study introduces three moderating variables for heterogeneity analysis: doctors' professional level (t), the Area a doctor located (t) (regional development level), and disease urgency level (t).

Doctors' professional level (t) (DUL_t) represents their professional qualifications and professional credibility on the platform, which may enhance patients' perceptions and acceptance of their behavior. Disease urgency level (t) (DUL_t) captures patients' sensitivity to response speed and service urgency. Patients with acute conditions may prefer faster and more efficient services, thus making the effect of doctor' s proactive behavior more pronounced. The Area a doctor located (t) ($Area_t$), or the level of regional development, affects the technological penetration of the platform and patients' adaptability to digital services. Differences in visibility and conversion efficiency of doctor behaviors may emerge across regions. These moderating variables are grounded in the logic of behavioral perception and service response matching, contributing to a deeper understanding of the main effect relationships.

(4) Control Variables

To enhance the explanatory power and identification validity of the model examining the relationship between doctors' proactive crafting behaviors and performance outcomes, this study controls for other potential confounding factors that might systematically affect performance outcomes. Four categories of control variables are included: latest review (t) (LR_t), online average price (t) (OAP_t), offline service prices (t) (OSP_t), and the total number of online and offline services (t) ($NOOS_t$).

Latest review (t) reflects the recent activity level and frequency of evaluations for a doctors' services, which may directly influence platform exposure and patient selection

preferences. Online average price (t) and Offline service prices (t) represent the pricing levels of the doctors' services. Price is an important reference factor in patient decision-making. The total number of online and offline services (t) measures the overall service intensity provided by the doctor, which may create scale effects or experience advantages affecting performance outcomes. Incorporating these control variables helps to isolate the effect of proactive crafting behavior on performance outcomes, thereby improving the robustness and credibility of the empirical analysis.

The definitions and measurement methods of variables are shown in Table 1.

Table 1. Summary of Variables.

Variable	Abbreviation	Variable Definition	Data Type
Online consultation volume (t+1)	OCV_{t+1}	The number of online consultations completed by doctors in the following month. A higher value indicates a greater volume of online consultations. Apply the natural log (ln) transformation to the value.	Continuous
Offline service volume (t+1)	OSV_{t+1}	The number of offline outpatient visits completed by doctors in the following month. A higher value indicates a greater volume of offline services. Apply the natural log (ln) transformation to the value.	Continuous
User evaluation performance (t+1)	UEP_{t+1}	The subjective rating given by platform patients to doctor services. The value ranges from 0 to 100; a score closer to 100 indicates higher patient satisfaction.	Continuous
Proactive crafting index (t)	PCI_t	The overall degree to which doctors actively shape their work behavior on the platform, calculated using the entropy weight method. The value ranges from 0 to 1; a value closer to 1 indicates a higher level of proactive crafting behavior.	Continuous
Latest Review (t)	LR_t	The latest number of comments from patients on the platform about the doctor. A higher value reflects a greater number of recent evaluations.	Continuous
Online average price (t)	OAP_t	The average price for doctors' online consultations. The value represents consultation pricing; a higher value indicates a higher online service fee.	Continuous
Offline service price (t)	OSP_t	Doctors' offline consultation prices. The value represents pricing for offline services; a higher value indicates a higher fee.	Continuous
Number of online and offline services (t)	$NOOS_t$	Total number of online and offline consultations by doctors. A higher value indicates a greater total number of services provided.	Continuous
Disease urgency level (t)	DUL_t	The urgency level of the disease consulted by the patient. Determined by whether the consultation was with an emergency department; acute = 1, non-acute = 0.	Binary (Nominal)
Doctors' professional level	DPL_t	Professional title hierarchy of doctors. Categorized by doctor rank; chief physicians are coded as high-level (1), all others as non-high-level (0).	Binary (Ordinal)
Area (t)	$Area_t$	Doctors' location. Categorized by region based on National Bureau of Statistics classifications: Eastern = 1, Western = 2, Central = 3, Northeastern = 4.	Categorical

4.3. Research Model

The research model of this study is shown in Figure 1.

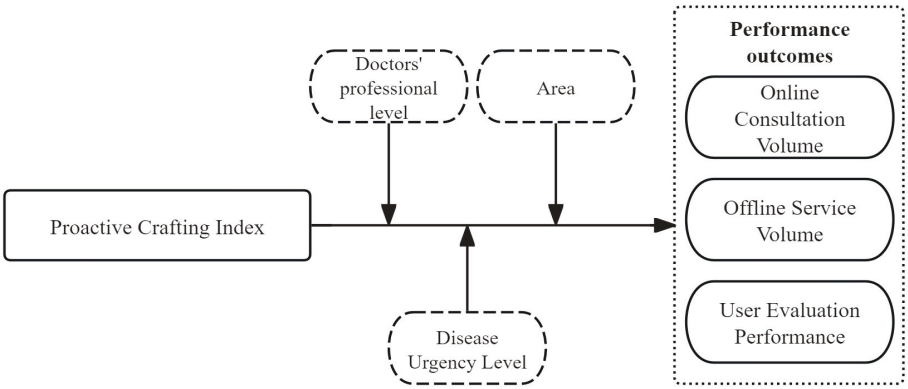


Figure 1. Research Model.

5. Results Analysis

5.1. Descriptive Statistics

The dataset includes over 2800 doctors from various regions, specialties, and professional levels across the platform. These doctors span both high-level and non-high-level titles and cover a broad range of clinical departments and geographic locations, which reflects the diversity of the platform's overall population. This study conducted descriptive statistics using Stata 17 (see Table 2). The results indicate that the distribution of the sample is highly representative. Both OCV_{t+1} and OSV_{t+1} exhibit considerable variation, reflecting differences in doctors' activity levels regarding patient consultations on the platform. The mean of UEP_{t+1} reaches 98.392, with a maximum value of 100, indicating that the majority of doctors received favorable patient evaluations, which reflects a notable "positive review bias" in the rating mechanism. Additionally, the maximum value of the $NOOS_t$ is as high as 98,797, far exceeding the median of 379, suggesting a highly skewed distribution in consultation volume. Overall, the data characteristics support the feasibility of conducting subsequent analyses.

Table 2. Descriptive Statistics.

VarName	Obs	Mean	SD	Min	Median	Max
OCV_{t+1}	23,455	1.602	1.346	0	1.386	5.659
OSV_{t+1}	10,969	1.985	1.499	0	1.792	5.591
UEP_{t+1}	42,500	98.392	3.311	75	99.700	100
PCI_t	30,102	0.243	0.120	0	0.287	1
LR_t	42,211	129.092	645.516	0	27	29,132
OAP_t	42,504	104.191	103.566	0	75	1333.333
OSP_t	42,504	27.240	16.337	4	25	100
$NOOS_t$	42,504	1778.830	4458.605	0	379	98,797
DUL_t	42,504	0.133	0.339	0	0	1
DPL_t	42,504	0.715	0.451	0	1	1
$Area_t$	42,504	13.294	12.027	1	11	29

Notes: This table is based on original panel data collected from www.wedoctor.com (accessed on 7 June 2023).

5.2. Correlation Analysis

Correlation analysis was conducted using Pearson correlation test in Stata 17 (see Table 3). The results show a significant positive correlation between doctors' proactive crafting behavior and online consultation volume, supporting the theoretical expectation that proactive crafting can enhance platform exposure and patients' willingness to consult. Although the correlation with OSV_{t+1} is negative, its absolute value is small, suggesting a potential nonlinear relationship. The PCI_t also exhibits a weak positive correlation with UEP_{t+1}; despite the limited correlation strength, the direction is consistent. Overall, all

correlation coefficients are below 0.8, indicating no serious multicollinearity issues among variables and providing a solid foundation for subsequent regression model estimation.

Table 3. Correlation Analysis.

	OCV_{t+1}	OSV_{t+1}	UEP_{t+1}	PCI_t	LR_t	OAP_t	OSP_t	$NOOS_t$	DPL_t	DUL_t	$Area_t$
OCV_{t+1}	1										
OSV_{t+1}	0.286 ***	1									
UEP_{t+1}	0.076 ***	0.018 *	1								
PCI_t	0.160 ***	−0.011 ***	0.054 ***	1							
LR_t	0.375 ***	0.215 ***	0.042 ***	0.236 ***	1						
OAP_t	0.051 ***	0.280 ***	0.024 ***	−0.062 ***	−0.001	1					
OSP_t	−0.043 ***	0.281 ***	0.010 **	−0.081 ***	−0.043 ***	0.485 ***	1				
$NOOS_t$	0.348 ***	0.464 ***	0.016 ***	0.114 ***	0.461 ***	0.258 ***	0.156 ***	1			
DPL_t	−0.063 ***	0.188 ***	0.029 ***	−0.085 ***	−0.042 ***	0.287 ***	0.451 ***	0.125 ***	1		
DUL_t	−0.015 **	−0.013	0.021 ***	0.014 **	0.006	0.030 ***	−0.012 **	−0.032 ***	−0.006	1	
$Area_t$	0.002	−0.254 ***	−0.020 ***	−0.031 ***	−0.008 *	−0.179 ***	−0.419 ***	−0.106 ***	−0.023 ***	−0.035 ***	1

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

5.3. Multicollinearity Test

Variance Inflation Factor (VIF) analysis was performed using Stata 17 to test for multicollinearity, and the results show that all explanatory variables have VIF values within a reasonable range (see Table 4), with a maximum value of 2.11 (OSP_t) and most variables below 1.5, indicating no serious multicollinearity problem in the regression model. The PCI_t has a VIF of 1.08, demonstrating good identification and independence as the main explanatory variable. The reciprocal of VIF values (1/VIF) all exceed 0.47, further confirming variable independence. These test results ensure the stability of parameter estimates in the subsequent multiple regression analyses.

Table 4. Multicollinearity Test.

Var	OCV_{t+1}	OSV_{t+1}	UEP_{t+1}	PCI_t	LR_t	OAP_t	OSP_t	$NOOS_t$	DPL_t	DUL_t	$Area_t$
VIF	1.32	1.36	1.06	1.08	1.25	1.37	2.11	1.56	1.32	1.01	1.52
1/VIF	0.757	0.735	0.947	0.929	0.803	0.728	0.474	0.641	0.757	0.992	0.658

5.4. Two-Way Fixed Effects Analysis

Based on the Hausman test conducted using Stata 17 ($p < 0.05$), this study adopts fixed effects modeling. The White test (also performed in Stata 17) indicates heteroscedasticity in the data; thus, robust standard errors are employed. Controlling for individual doctor and time fixed effects, the two-way fixed effects model results are presented in Table 5. After controlling for individual and temporal fixed effects, regression results show that the impact of proactive crafting behavior on online and offline performance exhibits significant nonlinear characteristics. Models (1) and (2) support the hypothesized inverted U-shaped relationship in H1 and the U-shaped relationship in H2, respectively.

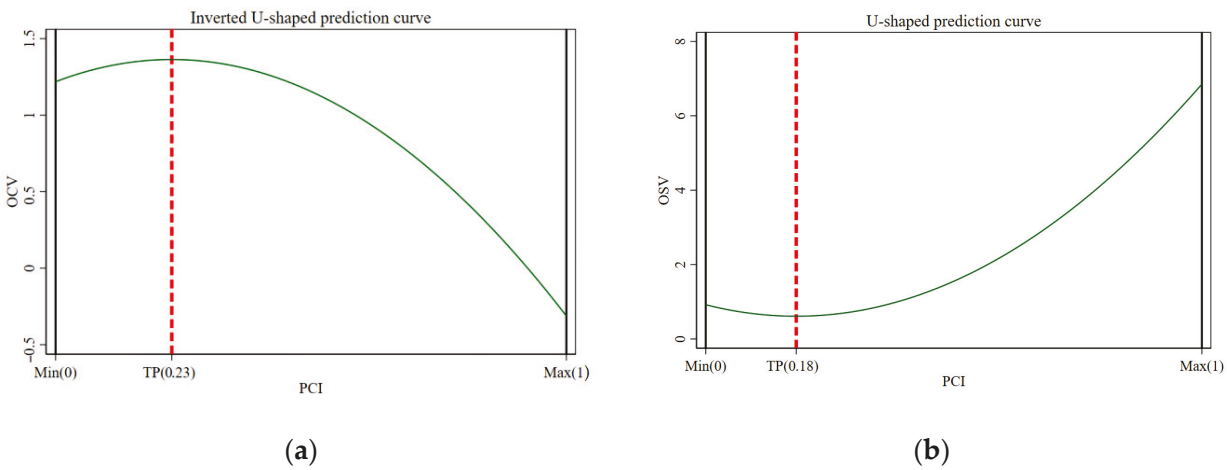
This study further applies the Utest (a formal test for U-shaped and inverted U-shaped nonlinear relationships) to validate these forms. And generate a prediction graph through Stata17, as shown in Figure 2. The results indicate that the online consultation model satisfies the unimodal property of an inverted U-shape, while the offline consultation model satisfies the unimodal property of a U-shape, both passing the Utest significance test. These findings not only confirm the theoretical nonlinear specifications but also statistically strengthen the credibility and identification of Hypotheses 1 and 2. The relatively low R^2 of Model (3) in predicting UEP_{t+1} may be attributed to the characteristics and measurement properties of the UEP data. Our results show that the mean UEP_{t+1} score is 98.39 (out of 100), with a median of 99.7, and over 90% of the evaluations fall within the 95–100 range. This strong clustering of high scores likely limits the model's ability to detect meaningful associations. Additionally, UEP_{t+1} may be influenced by unobserved factors

such as medical outcomes or individual user preferences. Nonetheless, despite the model's limited explanatory power, the results indicate that PCI_t has no significant effect on UEP_{t+1} (coefficient = 0.011, $p > 0.1$). This finding is consistent with the data characteristics: due to the heavy concentration of high scores, variations in proactive behavior may not translate into observable differences in user ratings, which still provides descriptive insights.

Table 5. Fixed effects analysis results.

	(1)	(2)	(3)
	OCV_{t+1}	OSV_{t+1}	UEP_{t+1}
PCI_t	1.279 **	−3.413 **	0.011
	−0.46	−1.351	−0.167
$PCI^2{}_t$	−2.809 **	9.341 **	
	−1.213	−3.831	
LR_t	0.001	0.000	0.000 **
	0.000	0.001	0.000
OAP_t	0.003 **	0.001	−0.001
	−0.001	−0.001	0
OSP_t	−0.017	0.009	0.003
	−0.012	−0.05	−0.008
$NOOS_t$	0.001 ***	0.001 ***	−0.001 **
	0.000	0.000	0.001
_cons	1.218 **	−0.077	98.754 ***
	−0.411	−1.647	−0.243
ID Fixed	Yes	Yes	Yes
Month Fixed	Yes	Yes	Yes
N	18,000	8827	30,000
R^2	0.064	0.443	0.004
R^2_a	0.064	0.442	0.004
F	97.673	404.094	4.843

Notes: ** $p < 0.05$, *** $p < 0.001$.



(**a**) (**b**)

Figure 2. Prediction curve: (**a**) OCV prediction curve; (**b**) OSV prediction curve.

5.5. Robustness Tests

5.5.1. Removal of Outliers

A robustness test was conducted by applying a 1% winsorization to the data using Stata 17, as shown in Table 6. To reduce the influence of extreme values on the estimation results, the main variables were winsorized at the 1% level. After this treatment, the study continued to use the two-way fixed effects model with robust standard errors for regression analysis. After this treatment, the regression results remained highly consistent

with the original models: OCV_{t+1} still exhibited a significant inverted U-shaped pattern, while OSV_{t+1} maintained a typical U-shaped trend. This indicates that extreme sample points have minimal interference with the overall trend, confirming the robustness of the regression findings.

Table 6. Robustness Test (1% Tail Shrinkage).

	OCV_{t+1}	OSV_{t+1}	UEP_{t+1}
PCI_t	1.466 ***	−1.719 **	−0.006
	−0.385	−0.739	−0.076
PCI^2_t	−3.348 ***	4.456 **	
	−0.986	−2.039	
LR_t	0.001	0.000	0.000 **
	0.000	0.001	0.000
OAP_t	0.002 **	0.001	0
	−0.001	−0.001	0
OSP_t	−0.016	0.008	0.007
	−0.012	−0.05	−0.008
$NOOS_t$	0.000 **	0.001 ***	−0.000 **
	0.001	0.001	0.001
_cons	1.375 ***	−0.122	98.657 ***
	−0.389	−1.633	−0.232
ID Fixed	Yes	Yes	Yes
Month Fixed	Yes	Yes	Yes
N	18,000	8827	30,000
R^2	0.063	0.451	0.008
R^2_a	0.062	0.451	0.008
F	100.639	420.567	6.728

Notes: ** $p < 0.05$, *** $p < 0.001$.

5.5.2. Control Variable Adjustment: Replacing LR with NNR

Building upon the original model, this study replaces the original control variable—LR_t—with Number of Negative Reviews (t) (NNR_t) using Stata 17 for analysis, considering that negative feedback on online platforms may have a more substantial impact on a doctor's reputation and performance. As explicit signals of service dissatisfaction, negative reviews may more effectively influence patients' decisions regarding future consultations or referrals, thereby offering a more precise control for the effects of patient perception on performance outcomes. Negative reviews, as explicit signals of service failure, may influence patients' decisions regarding revisit or referral. Therefore, incorporating this variable helps more comprehensively control for the effect of patients' subjective perceptions on performance outcomes, thereby improving the precision in identifying the net effect of proactive behaviors.

The results after inclusion (see Table 7) indicate that the nonlinear relationship between proactive crafting behavior and performance remains robust. The NNR_t variable itself shows a significant negative impact on User Evaluation Performance (UEP_{t+1}) but a significant positive impact on Offline Service Volume (OSV_{t+1}), confirming its explanatory role in the model. This suggests that while negative reviews may harm perceived service quality, they may also reflect higher patient engagement or exposure, thereby increasing the likelihood of subsequent offline visits.

Table 7. Robustness test (Adjusting control variables).

	OCV_{t+1}	OSV_{t+1}	UEP_{t+1}
PCI_t	1.473 ***	−1.292 *	−0.002
	−0.385	−0.676	−0.076
PCI^2_t	−3.375 ***	3.223 *	
	−0.985	−1.855	
NNR_t	0.015	0.037 **	−0.062 **
	−0.01	−0.014	−0.021
OAP_t	0.002 **	0.001	0
	−0.001	−0.001	0
OSP_t	−0.016	0.008	0.006
	−0.012	−0.05	−0.008
$NOOS_t$	0.000 **	0.001 ***	0.001
	0.000	0.000	0.000
_cons	1.378 ***	0.034	98.666 ***
	−0.387	−1.617	−0.238
ID Fixed	Yes	Yes	Yes
Month Fixed	Yes	Yes	Yes
N	18,000	8827	30,000
R^2	0.063	0.452	0.014
R^2_a	0.063	0.452	0.013
F	90.615	389.142	6.792

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

5.5.3. Recalculation of the Proactive Crafting Index

To further validate the rationality of the construction of the main explanatory variable, this study recalculated the PCI_t using the improved entropy weight method. Unlike the traditional entropy method, which relies solely on information entropy for indicator weighting, the improved entropy weight method incorporates standard deviation and range as additional metrics to adjust the weights. This enhancement increases sensitivity to differences in indicator distributions.

This optimization mechanism prevents disproportionate weighting caused by concentrated values or extreme outliers in any single indicator, thereby improving both the stability and theoretical expressiveness of the index. After introducing this new variable, the study conducted regression analysis using the two-way fixed effects model with robust standard errors via Stata 17. The re-estimated results (see Table 8) continue to confirm the inverted U-shaped relationship for Online Consultation Volume (OCV_{t+1}) and the U-shaped relationship for Offline Service Volume (OSV_{t+1}). These findings indicate that the construction method of the PCI_t does not materially affect the core conclusions, and demonstrates the robustness of the methodology in empirical application.

Table 8. Robustness Test (Improved Calculation Method).

	OCV_{t+1}	OSV_{t+1}	UEP_{t+1}
$NewPCI_t$	1.635 ***	−2.404 ***	−0.036
	−0.462	−0.44	−0.123
$NewPCI^2_t$	−6.797 **	12.568 ***	
	−2.316	−1.924	
LR_t	0.001	0.000	0.000 **
	0.000	0.001	0.000
OAP_t	0.002 **	0.001	0
	−0.001	−0.001	0

Table 8. *Cont.*

	OCV_{t+1}	OSV_{t+1}	UEP_{t+1}
OSP_t	−0.016	0.008	0.007
	−0.012	−0.05	−0.008
$NOOS_t$	0.000 ***	0.001 ***	−0.000 **
	0.000	−0.001	0.000
_cons	1.331 ***	0.175	98.660 ***
	−0.39	−1.633	−0.231
ID Fixed	Yes	Yes	Yes
Month Fixed	Yes	Yes	Yes
N	18,000	8826	30,000
R^2	0.063	0.454	0.008
R^2_a	0.063	0.453	0.008
F	100.359	440.954	6.724

Notes: ** $p < 0.05$, *** $p < 0.001$.

5.6. Endogeneity Test

Considering that doctors' proactive behaviors may be simultaneously influenced by feedback from performance outcomes (e.g., better performance leading to increased proactivity), this study conducted an endogeneity test to mitigate potential bias. Lagged values of the PCI_t and $PCI^2{}_t$ term were used as instrumental variables. The choice of lagged variables is based on the assumption that past behaviors are correlated with current proactive tendencies but are less likely to be directly affected by contemporaneous performance shocks, thus satisfying the relevance and exogeneity conditions for valid instruments [77–80]. In this study, the lagged variable (PCI_{t-1}) affects performance at time t+1 (e.g., OCV_{t+1}) indirectly by influencing current proactive behavior (PCI_t), rather than exerting a direct effect on future performance. This approach is grounded in the notion that patients' decisions to initiate online consultations or seek offline visits are predominantly driven by the doctor's current behavior. As a result, the lagged variable serves as a valid instrumental variable, being strongly correlated with the endogenous regressor (current proactive behavior) and, crucially, uncorrelated with the random disturbance term of the outcome equation, thereby satisfying the key conditions for an effective instrument. Moreover, the use of lagged independent variables has been validated in prior studies as an effective strategy to mitigate contemporaneous endogeneity [79,80].

The lagged variables were generated using Stata 17, followed by a test using the instrumental variable method. The results are shown in Table 9. The first-stage F-statistic was significant, indicating a strong correlation between the instruments and endogenous regressors, ruling out weak instrument problems. The CD Wald F statistic further verifies the validity of the instruments, showing that they can effectively explain the variation in the endogenous regressors. The SW S statistic test results demonstrate that the instruments satisfy the exogeneity condition, ensuring their validity in the model. Since the model is exactly identified, with the number of instruments equal to the number of endogenous regressors, no overidentification test (such as the Sargan test) is required. Therefore, based on these statistical test results, we can confirm that the selected instruments are valid and reliable.

Second-stage results showed that the lagged index maintains the inverted U-shaped relationship with OCV_{t+1}, consistent with the main model. The relationship with OSV_{t+1} remains U-shaped, also consistent with theoretical expectations. These findings suggest the instrument design is appropriate, effectively controlling for reverse causality and omitted variable bias, thereby further validating the reliability of the model results.

Table 9. Instrumental Variable Method.

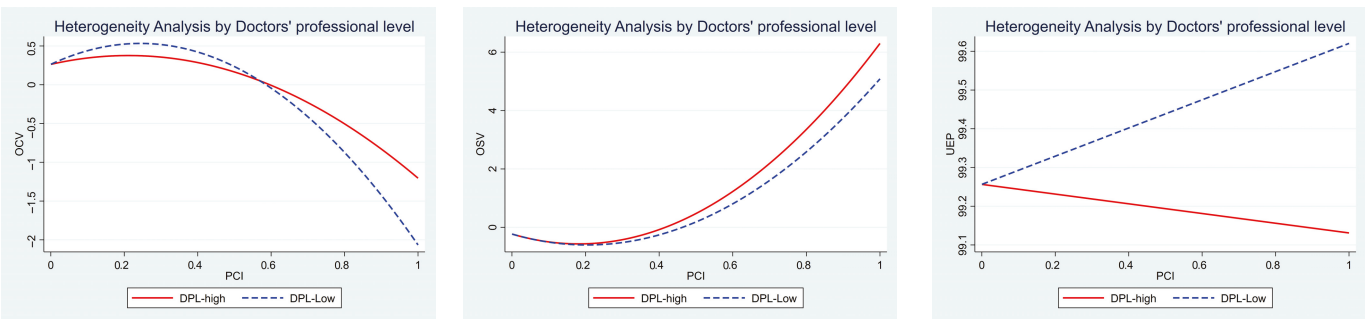
	PCI_t	$PCI^2{}_t$	OCV_{t+1}	PCI_t	$PCI^2{}_t$	OSV_{t+1}	PCI_t	UEP_{t+1}
$L.PCI_t$	−0.05 **	−0.088 ***		−0.111 **	−0.074 ***		0.222 ***	
	−0.033	−0.013		−0.055	−0.021		−0.009	
$L.PCI^2{}_t$	0.629 ***	0.437 ***		0.737 ***	0.379 ***			
	−0.083	−0.032		−0.143	−0.055			
PCI_t			5.443 **			−37.303 ***		−0.194
			−2.598			−13.344		−0.232
$PCI^2{}_t$			−13.955 **			93.726 ***		
			−5.696			−31.165		
LR_t	0.001	0.000	0.000 **	0.001	0.000	0.000 **	0.000	0.000 **
	0.000	0.001	0.000	0.000	0.001	0.000	0.001	0.000
OAP_t	0.000	−0.000 **	0.002 **	0.001	−0.000 **	0.003 *	0.001	0.001
	0.001	0.000	−0.001	0.000	0.001	−0.002	0.001	0.002
OSP_t	0.002 **	0.001	−0.015	0.004	0.002	0.019	−0.001	0.005
	−0.001	0.000	−0.013	−0.003	−0.001	−0.048	−0.001	−0.007
$NOOS_t$	0.000	0.000 ***	0.000 ***	0.001	0.000 ***	0.000	0.000 **	−0.000 **
	0.001	0.001	−0.001	0.000	−0.001	0.000	0.000	0.000
ID Fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	14,827	14,827	14,827	6000	6000	6000	16,154	16,154
R^2			0.006			0.635		0.001
F			11.455			20.08		1.644
CD Wald F	93.058			17.14			662.02	
SW S stat.	11.987			35.434			0.981	

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

5.7. Heterogeneity Analysis

5.7.1. Heterogeneity Analysis by Doctors' Professional Level

Doctors were divided into high-level and non-high-level groups for analysis using Stata 17 (see Table A1 in Appendix A). The prediction charts are shown in Figure 3. For high-level doctors, the relationship between the Proactive crafting index and OCV_{t+1} exhibits a more significant inverted U-shaped pattern, while the relationship with OSV_{t+1} shows a U-shaped trend. In contrast, although the OCV_{t+1} for non-high-level doctors is significant, the OSV_{t+1} is not. This suggests that high-level doctors, due to their greater authority and trustworthiness, are more capable of converting proactive behaviors on the platform into performance outcomes, supporting Hypothesis H4a.



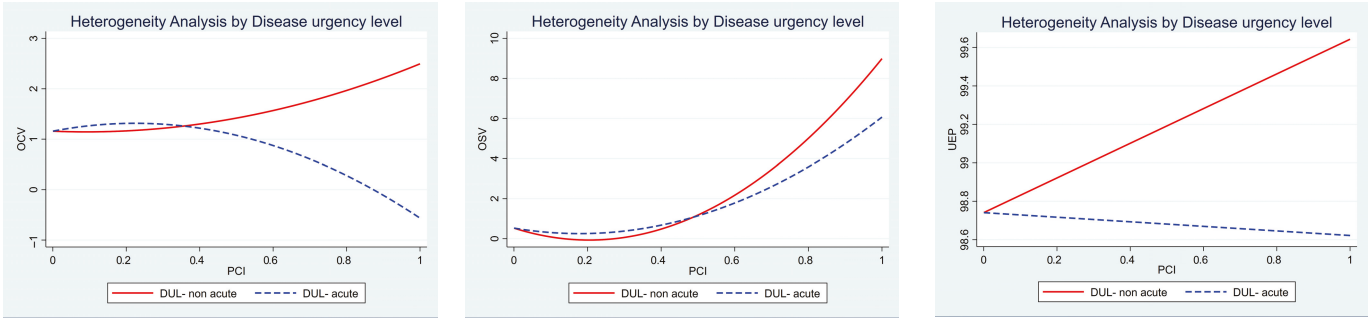
(**a**) (**b**) (**c**)

Figure 3. Heterogeneity Analysis by Doctors' professional level: (**a**) OCV prediction curve; (**b**) OSV prediction curve; (**c**) UEP prediction curve.

5.7.2. Heterogeneity Analysis by Disease Urgency Level

Disease types were categorized into non-acute and acute for analysis using Stata 17, with results shown in Table A2 in Appendix A. The prediction charts are shown in Figure 4.

Among acute patients, the inverted U-shaped relationship between doctor s' Proactive crafting index and OCV_{t+1} is significant. For non-acute patients, the U-shaped relationship with OSV_{t+1} is significant. Additionally, this study found that among non-acute patients, the proactive crafting index has a significant positive correlation with User evaluation performance. This may be because acute patients are more sensitive to response speed and service proactivity, enabling them to more quickly perceive doctors' enthusiasm, thus fostering trust and conversion behavior, which validates Hypothesis H4b.

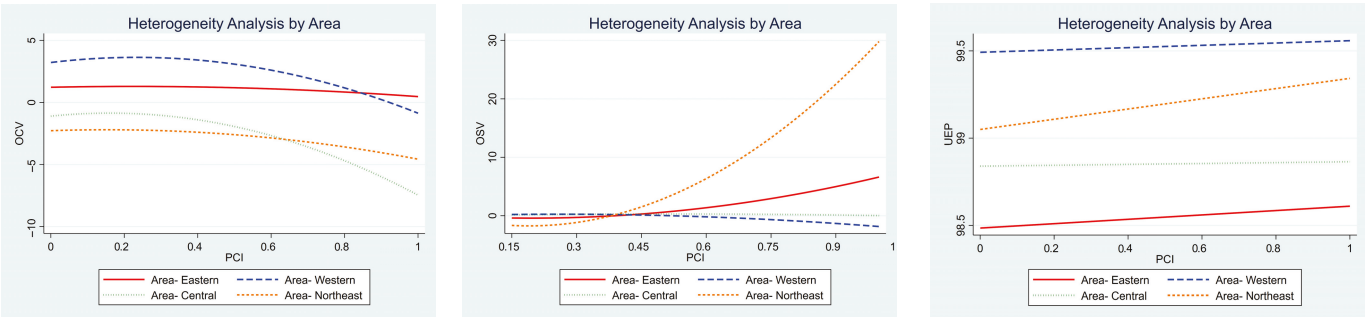


(**a**) (**b**) (**c**)

Figure 4. Heterogeneity Analysis by Disease urgency level: (**a**) OCV prediction curve; (**b**) OSV prediction curve; (**c**) UEP prediction curve.

5.7.3. Heterogeneity Analysis by Area

Regions were classified into four categories according to the National Bureau of Statistics standards: Eastern, Western, Central, and Northeast areas, and analyzed using Stata 17 accordingly (see Tables A3–A5 in Appendix A). The prediction charts are shown in Figure 5. The results indicate noticeable differences in the impact of doctors' proactive behavior on performance across regions. In the Eastern area, the inverted U-shaped relationship between proactive crafting and OCV_{t+1} is most pronounced, possibly due to higher platform usage frequency and greater patient digital literacy in this area. While certain trends exist in the Central and Western areas, the effects are relatively weaker. Regarding OSV_{t+1}, significant effects appear in both the Eastern and Northeast areas, whereas the Western and Central areas show no significant effects, which may relate to economic level differences or sample size disparities. UEP_{t+1} shows little variation across regions and remains statistically insignificant, likely due to the "high-score clustering" characteristic of rating data. Overall, these findings support Hypothesis H4c.



(**a**) (**b**) (**c**)

Figure 5. Heterogeneity Analysis by Area: (**a**) OCV prediction curve; (**b**) OSV prediction curve; (**c**) UEP prediction curve.

6. Discussion and Conclusions

6.1. Implications for Research

This study systematically investigates the impact of doctors' proactive crafting behaviors on multidimensional performance outcomes within online healthcare platforms. By incorporating nonlinear modeling and heterogeneity analysis, it extends the theoretical boundaries of healthcare operations management and platform behavior research. Compared to existing literature, this study contributes novel insights in terms of theoretical framing, methodological design, and empirical findings.

First, this study enriches the understanding of the relationship between doctors' proactive behavior and performance. Prior studies generally suggest that doctors' participation, activity level, and engagement on online platforms positively influence their performance [21–25], but most assume a linear relationship, implying that more effort yields better outcomes. In contrast, by constructing a Proactive Crafting Index (PCI), our empirical results reveal that proactive behavior exhibits an inverted U-shaped relationship with online performance and a U-shaped relationship with offline performance, highlighting the inherently nonlinear nature of these behavioral effects. These findings suggest that there is an optimal range of proactivity: moderate engagement enhances visibility and patient conversion, while excessively low or high levels may diminish performance returns. This aligns with the "boundary" in Job Crafting Theory, which posits that while proactive efforts can improve resource acquisition and job fit, they are subject to psychological and contextual constraints. As proactive behavior intensifies beyond a certain point, it may trigger fatigue, social strain, or resistance from others, thereby reducing its effectiveness [81–83]. In online healthcare settings, the boundary may be even more salient, as platform mechanisms and patient expectations jointly shape the 'acceptable' range of doctor engagement [12]. In our analysis, H3, which hypothesized a positive association between proactive crafting and user evaluation performance (UEP), yielded a statistically non-significant result. We attribute this mainly to the measurement limitations of the UEP variable, particularly its ceiling effect, with over 90% of observations clustered between 95 and 100. This rating inflation likely reduces score variation, limiting the model's ability to detect meaningful differences in proactive behavior. Therefore, the nonsignificant effect in H3 may stem from the limited variance and evaluative saturation of the dependent variable, rather than an absence of substantive impact.

Second, our results emphasize the potential adverse effects of excessive proactivity—an area often overlooked in the literature. In highly digitized, algorithm-driven online healthcare environments, doctors' behavioral frequency and response speed are tracked and factored into platform recommendation systems [38–40]. While proactive behaviors can enhance algorithmic ranking and increase doctors' exposure on online platforms, overly frequent accepting consultations may backfire—being perceived by patients as formulaic, commercially driven, or even manipulative. Such patterns may signal a lack of individualized attention, undermining the perceived authenticity of the doctor's communication. As patients in online healthcare environments often rely on subtle cues—such as message tone, response timing, and personalization—to infer service motivation, excessive behavioral frequency may be interpreted as strategic self-promotion rather than genuine care. Prior studies have shown that when users detect impression management or self-serving intent in digital service contexts, it can erode trust and reduce willingness to engage [84–86]. Studies have highlighted that patients are highly sensitive to doctors' service motivations, especially in the absence of face-to-face interaction. In such contexts, trust is built more through perceived sincerity, communication tone, and platform cues rather than traditional doctor-patient rapport [22,24]. Our findings support this "too much of a good thing" logic

and suggest that platform managers should help doctors find a balance between visibility and authenticity.

Third, through heterogeneity analysis, this study identifies the boundary conditions under which proactive behavior influences performance, thereby enriching the external validity of doctor behavior research. The results show that high-level doctors are more likely to translate proactive behavior into performance benefits, possibly due to the alignment between their professional authority, platform reputation, and patient expectations. For acute disease patients, responsiveness and enthusiasm are more highly valued, thus the marginal return on proactive behavior is greater. Regional analysis shows that in eastern regions, where digital adoption and platform usage are higher, proactive behavior has a stronger impact on performance, while in central and western regions, limited resources and information asymmetry weaken the effect. These findings underscore that the "conversion efficiency" of proactive behavior is not determined by behavior alone but is jointly shaped by individual attributes, patient types, and the surrounding platform environment. This represents a valuable supplement to prior studies that have not fully considered contextual differences [49,69].

In addition, this study advances the methodology of behavioral research on digital platforms. We construct the PCI using an Improved Entropy Weight Method, which, compared to the traditional entropy-based approach, incorporates standard deviation and range as auxiliary weighting factors. This increases the sensitivity of the index to distributional differences, enhancing its stability and theoretical clarity. Furthermore, to address potential endogeneity issues arising from reverse causality between behavior and performance, we adopt lagged values of the PCI and its squared term as instrumental variables. The results validate both the relevance and exogeneity of the instruments, confirming the robustness of our model.

In conclusion, this study supports the general notion in existing literature that proactive participation by doctors is valuable, but further reveals that this impact is nonlinear and context-dependent. By developing a multidimensional performance framework, identifying nonlinear patterns, and introducing multilevel moderators, this research presents a more nuanced and explanatory model of the behavior–performance relationship on online healthcare platforms. It not only extends the theoretical application of Job Crafting Theory in digital healthcare settings but also offers practical implications for guiding and incentivizing doctor behavior on such platforms.

6.2. Implications for Practice

This study also offers several important managerial implications for doctors, patients, and operators of online healthcare platforms. To begin with, our findings demonstrate that doctors' proactive crafting behaviors can significantly influence their online and offline performance outcomes, but the effect is nonlinear. This suggests that platforms should encourage doctors to maintain a moderate level of proactive engagement, where active service and interaction improve visibility and patient attraction, while avoiding excessive activity that may appear mechanical or profit-driven and negatively affect patient trust. Second, the identification of inverted U-shaped and U-shaped relationships between proactive behavior and performance provides actionable insights for platform algorithm design and doctor performance management. Platform operators could integrate the findings of this study into their recommendation systems by setting thresholds for proactive behaviors. For example, doctors displaying excessively high proactive behavior may be flagged for needing balance to prevent burnout or customer fatigue. Similarly, platforms could adjust incentive structures to reward moderate proactivity, as research shows that this level enhances patient trust and engagement. Third, the heterogeneity findings suggest that

platform operators should adopt differentiated operational strategies based on doctors' characteristics and service contexts. For example, high-level doctors are more effective in converting proactive behavior into performance, particularly in offline services. Therefore, platforms can prioritize promoting these doctors for high-urgency or complex cases to improve patient outcomes and platform credibility. Conversely, lower-level doctors may benefit more from support programs that guide the quality rather than quantity of engagement to build trust and improve long-term performance.

Moreover, the results show that patients with acute conditions respond more positively to proactive behaviors, particularly in online consultations. Platforms could design patient–doctor matching systems that consider disease urgency, thereby assigning more responsive doctors to patients requiring fast and effective interaction. This targeted matching can enhance both service satisfaction and medical outcomes. Additionally, the regional heterogeneity observed in our study implies that platform operators in less-developed areas (e.g., central and western regions) may need to invest more in patient education and digital literacy to help users understand and value doctors' proactive services. Meanwhile, platforms in digitally mature areas like eastern China can leverage behavioral data more effectively for real-time performance feedback and service personalization. Finally, the findings underscore the need for dynamic, multidimensional performance evaluation systems. Platforms should move beyond static metrics such as consultation volume and incorporate behavioral quality indicators (e.g., response speed, communication tone), patient feedback, and conversion rates. Such systems can provide more holistic and fair assessments of doctors' contributions and foster sustainable service motivation.

6.3. Limitations and Future Research

While this study provides meaningful theoretical and empirical insights, several limitations should be noted, offering valuable directions for future research. The analysis is based on data from a single online healthcare platform in China. Although the platform is representative in terms of user base and service structure, the observed relationships—especially those derived from the Proactive Crafting Index (PCI)—may be shaped by platform-specific incentive mechanisms, sociocultural expectations, or regulatory environments. Future research could extend the scope by incorporating data from multiple platforms or conducting cross-country comparative studies to examine whether the observed nonlinear and heterogeneous effects are consistent across different healthcare systems. Another limitation lies in the focus on doctor-side behaviors, without incorporating the dynamic interactions between doctors and patients. In real-world platform environments, patient responses—such as feedback, follow-up behavior, and trust formation—can significantly influence doctors' behavioral decisions, forming a recursive feedback loop. Future studies could explore these bilateral dynamics by analyzing interactional data and examining how mutual adaptation between doctors and patients affects performance outcomes. While this study adopts a multidimensional framework to measure performance—including online consultation volume, offline service volume, and user evaluation performance—some critical dimensions may still be overlooked. For instance, clinical effectiveness, patient retention, or long-term health improvements are also essential outcomes in healthcare services. Future research could incorporate clinical indicators or third-party data to provide a more comprehensive assessment of performance. Additionally, this study does not explicitly address potential biases arising from platform algorithm opacity or patient self-selection. The recommendation and exposure mechanisms on digital health platforms are often governed by proprietary algorithms that are not fully disclosed to users or researchers. Such opacity may introduce systematic bias in how doctors' behaviors are rewarded or penalized. Furthermore, patient choices—such as selecting doctors based on perceived popularity or availability—may lead

to selection effects that confound the relationship between doctor behavior and performance. Future studies should account for these potential biases, possibly through platform collaboration or experimental designs. Although methodological efforts were made to enhance robustness—including the application of the Improved Entropy Weight Method and the use of lagged instrumental variables to address endogeneity—there remains the possibility of unobserved confounding factors or biases introduced by opaque platform algorithms. Future studies could strengthen causal inference by employing experimental or quasi-experimental designs, such as leveraging natural experiments or platform-level policy changes. The UEP variable used in this study has a ceiling effect, as most ratings cluster in the higher range, limiting its ability to measure subtle differences in service quality. Future research could consider incorporating more granular patient feedback, such as satisfaction surveys, to complement this measurement limitation and capture finer variations in service quality. In addition, our current modeling approach treats user behavior and feedback in entropy weight method, which, while aligned with data availability and study scope, does not fully capture the complexity and heterogeneity of patient decision-making. Future research may explore this dimension more deeply by incorporating user-level behavioral modeling, segmentation strategies, or qualitative methods to uncover how different user types perceive and respond to doctors' proactive behaviors in platform-based healthcare environments.

In sum, future research should aim to expand the empirical scope, incorporate dynamic and bilateral interactions, and explore system-level governance factors in order to construct a more comprehensive understanding of proactive behavior and performance in online healthcare contexts.

Author Contributions: Conceptualization, W.L. and Y.Y.; methodology, Y.Y. and Z.B.; formal analysis, Y.Y. and S.S.; data curation, W.L.; writing—original draft preparation, W.L. and Y.Y.; writing—review and editing, W.L. and Y.Y.; funding acquisition, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by "the Fundamental Research Funds for the Central Universities", NO. ND2024001.

Institutional Review Board Statement: Not applicable. Due to the nature of the study, Ethics Committee approval was not required.

Informed Consent Statement: Not applicable. We would like to clarify that our study is based on publicly available online consultation data obtained from a widely used Chinese online healthcare platform, with all data processing conducted in compliance with the platform's terms of service and relevant ethical guidelines. Therefore, informed consent was not required.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

OCV_{t+1}	Online consultation volume (t+1)
OSV_{t+1}	Offline service volume (t+1)
UEP_{t+1}	User evaluation performance (t+1)
PCI_t	Proactive crafting index (t)
LR_t	Latest Review (t)
OAP_t	Online average price (t)
OSP_t	Offline service prices (t)

$NOOS_t$	Number of online and offline services (t)
DUL_t	Disease urgency level (t)
DPL_t	Doctors' professional level (t)
$Area_t$	Area (t)
NNR_t	Number of Negative Reviews (t)

Appendix A

Table A1. Heterogeneity Analysis (DPL).

	OCV_{t+1}		OSV_{t+1}		UEP_{t+1}	
	DPL-high	DPL-Low	DPL-high	DPL-Low	DPL-high	DPL-Low
PCI_t	1.198 **	2.541 ***	−2.020 **	−2.412	−0.007	−0.011
	−0.409	−0.763	−0.853	−2.496	−0.078	−0.189
$PCI^2{}_t$	−2.890 **	−5.528 **	5.359 **	5.312		
	−1.049	−1.95	−2.362	−7.026		
LR_t	0.000	0.000	0.001	−0.001	0.000 **	0.000 *
	0.000	0.000	0	−0.001	0.000	0.000
OAP_t	0.002 *	0.005 *	0.001	0.006	0.000	0.000
	−0.001	−0.002	−0.001	−0.006	0.000	−0.001
OSP_t	−0.013	0.068	−0.044	0.000	0.018	−0.002
	−0.021	−0.089	−0.05	−0.001	−0.022	−0.002
$NOOS_t$	0.000 **	0.000 **	0.000 ***	0.001 ***	0	−0.000 **
	0.000	0.000	0.000	0.000	0.000	0.000
_cons	1.111	0.325	2.724	−0.186	98.024 ***	99.306 ***
	−0.77	−1.424	−1.78	−0.467	−0.707	−0.102
ID Fixed	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed	Yes	Yes	Yes	Yes	Yes	Yes
N	13,000	4965	7405	1422	22,000	8302
R^2	0.067	0.115	0.497	0.321	0.006	0.013
R^2_a	0.067	−0.379	0.496	0.317	0.006	0.012
F	74.176	45.982	440.212	21.308	5.545	2.818

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table A2. Heterogeneity Analysis (DUL).

	OCV_{t+1}		OSV_{t+1}		UEP_{t+1}	
	DUL-non acute	DUL-acute	DUL-non acute	DUL-acute	DUL-non acute	DUL-acute
PCI_t	0.304	1.545 ***	−4.911 **	−1.314	0.376 *	−0.062
	−1.25	−0.408	−1.831	−0.8	−0.225	−0.081
$PCI^2{}_t$	−0.055	−3.554 ***	11.574 **	3.531		
	−3.334	−1.04	−5.016	−2.208		
LR_t	0.000	0.000	0.002	0.000	0.000	0.000 **
	0.000	0.000	−0.002	0.000	0.000	0.000
OAP_t	0.001	0.003 **	0.003 **	0.000	0.000	0.000
	−0.002	−0.001	−0.002	−0.002	0.000	0.000
OSP_t	0.035	−0.017	0.146 ***	−0.015	−0.009	0.007
	−0.058	−0.012	−0.01	−0.054	−0.011	−0.009
$NOOS_t$	0.000 *	0.000 **	−0.001 **	0.001 ***	0.000	−0.000 **
	0.000	0.000	0.000	0.0000	0.000	0.000
_cons	−0.248	1.348 ***	0.2	0.471	99.118 ***	98.644 ***
	−1.692	−0.402	−1.232	−1.758	−0.308	−0.247
ID Fixed	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed	Yes	Yes	Yes	Yes	Yes	Yes
N	2357	16,000	1094	7733	3915	26,000
R^2	0.063	0.065	0.513	0.45	0.006	0.009
R^2_a	0.06	0.064	0.509	0.449	0.003	0.009
F	11.21	90.526	.	365.699	2.735	6.211

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table A3. OCV Heterogeneity Analysis (Area).

	OCV_{t+1}			
	Area-Eastern	Area-Western	Area-Central	Area-Northeast
PCI_t	0.906 **	3.753 **	2.646 *	0.998
	−0.413	−1.594	−1.401	−2.325
PCI^2_t	−2.201 **	−8.266 **	−8.368 **	−3.273
	−1.052	−4.003	−3.792	−6.125
LR_t	0.000	0.000	−0.001	0.001 *
	0.000	0.000	−0.001	0.000
OAP_t	0.002 **	−0.012 ***	0.001	0.013
	−0.001	−0.003	−0.002	−0.01
OSP_t	−0.009	−0.057	−0.025	0.153 ***
	−0.018	−0.064	−0.018	−0.02
$NOOS_t$	0.000 **	0.000 *	0.001 **	0.001
	0.000	0.000	0.000	0.000
_cons	1.267 *	3.027 ***	−0.051	−2.042 **
	−0.651	−0.861	−0.692	−1.006
ID Fixed	Yes	Yes	Yes	Yes
Month Fixed	Yes	Yes	Yes	Yes
N	12,000	1159	2101	602
R^2	0.047	0.097	0.065	0.097
R^2_a	0.046	0.09	0.061	0.084
F	55.985	9.68	10.343	.

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table A4. OSV Heterogeneity Analysis (Area).

	OSV_{t+1}			
	Area-Eastern	Area-Western	Area-Central	Area-Northeast
PCI_t	−2.070 **	2.171	1.395	−18.281 ***
	−0.766	−4.044	−1.69	−4.167
PCI^2_t	4.732 **	−4.266	−2.222	48.097 ***
	−2.071	−9.601	−4.917	−11.991
LR_t	0.001	0.000	0.003 **	−0.001
	−0.001	−0.001	−0.001	−0.009
OAP_t	−0.001	0.010 **	0.008	−22.033 **
	−0.001	−0.004	−0.007	−10.031
OSP_t	0.038	0.000	−0.108 **	0.000
	−0.067	0.000	−0.041	0.000
$NOOS_t$	0.001 ***	0.001 ***	−0.002 ***	0.014 ***
	0.000	0.000	0.000	−0.004
_cons	−1.431	−0.569	8.282 ***	1975.360 **
	−2.59	−0.545	−1.357	−900.741
ID Fixed	Yes	Yes	Yes	Yes
Month Fixed	Yes	Yes	Yes	Yes
N	6053	370	1207	157
R^2	0.474	0.438	0.579	0.469
R^2_a	0.474	0.426	0.575	0.44
F	326.709	23.779	107.349	.

Notes: ** $p < 0.05$, *** $p < 0.001$.

Table A5. UEP Heterogeneity Analysis (Area).

	UEP_{t+1}			
	Area-Eastern	Area-Western	Area-Central	Area-Northeast
PCI_t	−0.013	0.071	0.025	0.293
	−0.091	−0.072	−0.143	−0.223
LR_t	0.000	0.000 **	0.000 **	0.000
	0.001	0.000	0.000	0.001
OAP_t	0.000	0.000	−0.002	0.000
	0.000	0.000	−0.002	−0.002
OSP_t	0.013	−0.032	−0.006 **	0.012
	−0.012	−0.023	−0.003	−0.009
$NOOS_t$	0.000	−0.000 **	−0.001 **	−0.000 *
	0.000	0.001	0.001	0.001
_cons	98.383 ***	99.499 ***	98.844 ***	99.058 ***
	−0.417	−0.286	−0.232	−0.196
ID Fixed	Yes	Yes	Yes	Yes
Month Fixed	Yes	Yes	Yes	Yes
N	20,000	1937	3476	982
R^2	0.006	0.004	0.015	0.022
R^2_a	0.005	−0.001	0.012	0.012
F	3.02	1.16	2.651	2.515

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

References

1. Guo, S.; Guo, X.; Fang, Y.; Vogel, D. How Doctors Gain Social and Economic Returns in Online Health-Care Communities: A Professional Capital Perspective. *J. Manag. Inf. Syst.* **2017**, *34*, 487–519. [CrossRef]
2. Peng, Y.; Yin, P.; Deng, Z.; Wang, R. Patient–Physician Interaction and Trust in Online Health Community: The Role of Perceived Usefulness of Health Information and Services. *Int. J. Environ. Res. Public Health* **2020**, *17*, 139. [CrossRef]
3. Peng, M.; Zhu, K.; Gu, Y.; Yang, X.; Su, K.; Gu, D. How to Self-Disclose? The Impact of Patients' Linguistic Features on Doctors' Service Quality in Online Health Communities. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 56. [CrossRef]
4. Li, M.; Gu, D.; Li, R.; Gu, Y.; Liu, H.; Su, K.; Wang, X.; Zhang, G. The Impact of Linguistic Signals on Cognitive Change in Support Seekers in Online Mental Health Communities: Text Analysis and Empirical Study. *J. Med. Internet Res.* **2025**, *27*, e60292. [CrossRef]
5. Liu, H.; Zhang, Y.; Li, Y.; Albright, K. Better Interaction Performance Attracts More Chronic Patients? Evidence from an Online Health Platform. *Inf. Process Manag.* **2023**, *60*, 103413. [CrossRef]
6. Wang, L.; Yan, L.; Zhou, T.; Guo, X.; Heim, G.R. Understanding physicians' online-offline behavior dynamics: An empirical study. *Inf. Syst. Res.* **2020**, *31*, 537–555. [CrossRef]
7. Wrzesniewski, A.; Dutton, J.E. Crafting a Job: Revisioning Employees as Active Crafters of Their Work. *Acad. Manag. Rev.* **2001**, *26*, 179–201. [CrossRef]
8. Tims, M.; Bakker, A.B. Job Crafting: Towards a New Model of Individual Job Redesign. *SA J. Ind. Psychol.* **2010**, *36*, 1–9. [CrossRef]
9. Seppälä, P.; Hakanen, J.J.; Virkkala, J.; Tolvanen, A.; Punakallio, A.; Rivinoja, T.; Uusitalo, A. Can Job Crafting eLearning Intervention Boost Job Crafting and Work Engagement, and Increase Heart Rate Variability? Testing a Health Enhancement Process. *J. Occup. Health Psychol.* **2023**, *28*, 395–410. [CrossRef]
10. Yan, J.; Liang, C.; Gu, D.; Zhu, K.; Zhou, P. Understanding Patients' Doctor Choice Behavior: Elaboration Likelihood Perspective. *Inf. Dev.* 2024; *forthcoming*. [CrossRef]
11. Tims, M.; Bakker, A.B.; Derks, D. Development and validation of the job crafting scale. *J. Vocat. Behav.* **2012**, *80*, 173–186. [CrossRef]
12. Tims, M.; Bakker, A.B.; Derks, D. The Impact of Job Crafting on Job Demands, Job Resources, and Well-Being. *J. Occup. Health Psychol.* **2013**, *18*, 230–240. [CrossRef]
13. Demerouti, E. Design Your Own Job through Job Crafting. *Eur. Psychol.* **2014**, *19*, 237–247. [CrossRef]
14. Petrou, P.; Demerouti, E.; Peeters, M.C.W.; Schaufeli, W.B.; Hetland, J. Crafting a Job on a Daily Basis: Contextual Correlates and the Link to Work Engagement. *J. Organ. Behav.* **2012**, *33*, 1120–1141. [CrossRef]
15. Lu, C.Q.; Wang, H.J.; Lu, J.J.; Du, D.Y.; Bakker, A.B. Does Work Engagement Increase Person–Job Fit? The Role of Job Crafting and Job Insecurity. *J. Vocat. Behav.* **2014**, *84*, 142–152. [CrossRef]

16. Nagarajan, R.; Swamy, R.A.; Reio, T.G.; Elangovan, R.; Parayitam, S. The COVID-19 Impact on Employee Performance and Satisfaction: A Moderated Moderated-Mediation Conditional Model of Job Crafting and Employee Engagement. *Hum. Resour. Dev. Int.* **2022**, *25*, 600–630. [CrossRef]
17. Lichtenthaler, P.W.; Fischbach, A. The Conceptualization and Measurement of Job Crafting: A Critical Review. *Work Stress* **2019**, *33*, 147–172. [CrossRef]
18. Vogt, K.; Hakanen, J.J.; Brauchli, R.; Jenny, G.J.; Bauer, G.F. The Consequences of Job Crafting: A Three-Wave Study. *Eur. J. Work Organ. Psychol.* **2016**, *25*, 353–362. [CrossRef]
19. Shi, Y.; Xie, J.; Wang, Y.; Zhang, N. Digital job crafting and its positive impact on job performance: The perspective of individual-task-technology fit. *Adv. Psychol. Sci.* **2023**, *31*, 1133–1145. [CrossRef]
20. Liu, W.; Chang, W.; Guo, J.; Chen, S.; Wang, H.J. Electronic performance monitoring and job crafting: A psychological ownership perspective. *J. Digit. Manag.* **2025**, *1*, 9. [CrossRef]
21. Bakker, A.B.; Demerouti, E. Job Demands–Resources Theory: Taking Stock and Looking Forward. *J. Occup. Health Psychol.* **2017**, *22*, 273–285. [CrossRef]
22. Cao, Q.; Lu, Y.; Dong, D.; Tang, Z.; Li, Y. The Roles of Bridging and Bonding in Social Media Communities. *J. Am. Soc. Inf. Sci. Technol.* **2013**, *64*, 1671–1681. [CrossRef]
23. Yan, Z.; Wang, T.; Chen, Y.; Zhang, H. Knowledge Sharing in Online Health Communities: A Social Exchange Theory Perspective. *Inf. Manag.* **2016**, *53*, 643–653. [CrossRef]
24. Deja, M.; Hommel, M.; Weber-Carstens, S.; Moss, M.; von Dossow, V.; Sander, M.; Pille, C.; Spies, C. Evidence-Based Therapy of Severe Acute Respiratory Distress Syndrome: An Algorithm-Guided Approach. *J. Int. Med. Res.* **2008**, *36*, 211–221. [CrossRef]
25. Shah, A.M.; Naqvi, R.A.; Jeong, O.R. The Impact of Signals Transmission on Patients' Choice through E-consultation Websites: An Econometric Analysis of Secondary Datasets. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5192. [CrossRef] [PubMed]
26. Tan, H.; Yan, M. Physician-User Interaction and Users' Perceived Service Quality: Evidence from Chinese Mobile Healthcare Consultation. *Inf. Technol. People* **2020**, *33*, 1403–1426. [CrossRef]
27. Lu, Y.; Wang, Q. Doctors' Preferences in the Selection of Patients in Online Medical Consultations: An Empirical Study with Doctor–Patient Consultation Data. *Healthcare* **2022**, *10*, 1435. [CrossRef] [PubMed]
28. Liu, J.; Bian, Y.; Ye, Q.; Jing, D. Free for Caring? The Effect of Offering Free Online Medical-Consulting Services on Physician Performance in e-Health Care. *Telemed. E-Health* **2019**, *25*, 979–986. [CrossRef] [PubMed]
29. Zhang, Y.; Zhang, J. Spillover Effects of Physicians' Prosocial Behavior: The Role of Knowledge Sharing in Enhancing Paid Consultations Across Healthcare Networks. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 87. [CrossRef]
30. Lai, A.Y.; Wee, K.Z.; Frimpong, J.A. Proactive behaviors and health care workers: A systematic review. *Health Care Manag. Rev.* **2024**, *49*, 239–251. [CrossRef]
31. Ebert, T.; Bipp, T.; Debus, M.E. Stability, reciprocity, and antecedent-outcome relations of different job crafting forms. *Occup. Health Sci.* 2025; *forthcoming*. [CrossRef]
32. Topa, G.; Aranda-Carmena, M. Job crafting in nursing: Mediation between work engagement and job performance in a multisample study. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12711. [CrossRef] [PubMed]
33. Liu, S.; Si, G.; Gao, B. Which Voice are You Satisfied with? Understanding the Physician–Patient Voice Interactions on Online Health Platforms. *Decis. Support Syst.* **2022**, *157*, 113754. [CrossRef]
34. Donabedian, A. Evaluating the Quality of Medical Care. *Milbank Q.* **1966**, *44*, 166–206. [CrossRef]
35. Campbell, S.M.; Roland, M.O.; Buetow, S.A. Defining Quality of Care. *Soc. Sci. Med.* **2000**, *51*, 1611–1625. [CrossRef]
36. Goh, J.M.; Gao, G.; Agarwal, R. The Creation of Social Value: Can an Online Health Community Reduce Rural–Urban Health Disparities? *Manag. Inf. Syst. Q.* **2016**, *40*, 247–263. [CrossRef]
37. Chen, J.; Zhang, C.; Xu, Y. The Role of Mutual Trust in Building Members' Loyalty to a C2C Platform Provider. *Int. J. Electron. Commer.* **2020**, *24*, 206–231. [CrossRef]
38. Chen, Y.; Lee, S. User-Generated Physician Ratings and Their Effects on Patients' Physician Choices: Evidence from Yelp. *J. Mark.* **2024**, *88*, 77–96. [CrossRef]
39. Chen, L.; Tang, H.; Guo, Y. Effect of Patient-Centered Communication on Physician-Patient Conflicts from the Physicians' Perspective: A Moderated Mediation Model. *J. Health Commun.* **2022**, *27*, 164–172. [CrossRef]
40. Aalto, A.M.; Elovainio, M.; Tynkkynen, L.K.; Reissell, E.; Vehko, T.; Chydenius, M.; Sinervo, T. What Patients Think about Choice in Healthcare? A Study on Primary Care Services in Finland. *Scand. J. Public Health* **2018**, *46*, 463–470. [CrossRef]
41. Zhang, X.; Guo, X.; Lai, K.; Guo, F.; Li, C. Understanding Gender Differences in m-Health Adoption: A Modified Theory of Reasoned Action Model. *Telemed. E-Health* **2014**, *20*, 39–46. [CrossRef]
42. Gartner, J.B.; Lemaire, C. Dimensions of performance and related key performance indicators addressed in healthcare organisations: A literature review. *Int. J. Health Plan. Manag.* **2022**, *37*, 1941–1952. [CrossRef]

43. Carini, E.; Gabutti, I.; Frisicale, E.M.; Di Pilla, A.; Pezzullo, A.M.; de Waure, C.; Cicchetti, A.; Boccia, S.; Specchia, M.L. Assessing hospital performance indicators. What dimensions? Evidence from an umbrella review. *BMC Health Serv. Res.* **2020**, *20*, 1038. [CrossRef]
44. Liu, X.; Guo, H.; Wang, L.; Hu, M.; Wei, Y.; Liu, F.; Wang, X. Effect of prosocial behaviors on e-consultations in a web-based health care community: Panel data analysis. *J. Med. Internet Res.* **2024**, *26*, e52646. [CrossRef]
45. van Elten, H.J.; van der Kolk, B. Performance management, metric quality, and trust: Survey evidence from healthcare organizations. *Br. Account. Rev.* **2024**, 101511. [CrossRef]
46. Huang, Z.; Duan, C.; Yang, Y.; Khanal, R. Online Selection of a Physician by Patients: The Impression Formation Perspective. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 193. [CrossRef] [PubMed]
47. Wu, H.; Lu, N. Service Provision, Pricing, and Patient Satisfaction in Online Health Communities. *Int. J. Med. Inform.* **2018**, *110*, 77–89. [CrossRef]
48. Liu, X.; Guo, X.; Wu, H.; Wu, T. The Impact of Individual and Organizational Reputation on Physicians' Appointments Online. *Int. J. Electron. Commer.* **2016**, *20*, 551–577. [CrossRef]
49. Zhang, F.; Parker, S.K. Reorienting Job Crafting Research: A Hierarchical Structure of Job Crafting Concepts and Integrative Review. *J. Organ. Behav.* **2019**, *40*, 126–146. [CrossRef]
50. Giæver, F.; Løvseth, L.T. Exploring Presenteeism among Hospital Physicians through the Perspective of Job Crafting. *Qual. Res. Organ. Manag.* **2020**, *15*, 296–314. [CrossRef]
51. Nguyen, N.X.; Tran, K.; Nguyen, T.A. Impact of Service Quality on In-patients' Satisfaction, Perceived Value, and Customer Loyalty: A Mixed-Methods Study from a Developing Country. *Patient Prefer. Adherence* **2021**, *17*, 2523–2538. [CrossRef]
52. Guo, X.; Guo, S.; Vogel, D.; Li, Y. Online Healthcare Community Interaction Dynamics. *J. Manag. Sci. Eng.* **2016**, *1*, 58–74. [CrossRef]
53. He, Y.; Guo, X.; Wu, T.; Vogel, D. The Effect of Interactive Factors on Online Health Consultation Review Deviation: An Empirical Investigation. *Int. J. Med. Inform.* **2022**, *163*, 104781. [CrossRef]
54. Liu, X.; Chi, X.; Li, J.; Zhou, S.; Cheng, Y. Doctors' Self-Presentation Strategies and the Effects on Patient Selection in Psychiatric Department from an Online Medical Platform: A Combined Perspective of Impression Management and Information Integration. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 13. [CrossRef]
55. Igarashi, H.; Kitamura, T.; Ohuchi, K.; Mitoma, H. Consultation frequency and perceived consultation time in a Japanese psychiatric clinic: Their relationship with patient consultation satisfaction and depression and anxiety. *Psychiatry Clin. Neurosci.* **2008**, *62*, 129–134. [CrossRef]
56. Jing, L.; Shan, W.; Evans, R.; Shi, X. To Continue Consultation or Not? How Physicians' Information Cues Affect Patients' Continued Online Consultation Behavior. *Electron. Commer. Res.* 2024; *forthcoming*. [CrossRef]
57. Cao, X.; Liu, Y.; Zhu, Z.; Hu, J.; Chen, X. Online Selection of a Physician by Patients: Empirical Study from Elaboration Likelihood Perspective. *Comput. Hum. Behav.* **2017**, *73*, 403–412. [CrossRef]
58. Guetz, B.; Bidmon, S. The Impact of Social Influence on the Intention to Use Physician Rating Websites: Moderated Mediation Analysis Using a Mixed Methods Approach. *J. Med. Internet Res.* **2022**, *24*, e37505. [CrossRef]
59. Sun, Q.; Tang, G.; Xu, W.; Zhang, S. Social media stethoscope: Unraveling how doctors' social media behavior affects patient adherence and treatment outcome. *Front. Public Health* **2024**, *12*, 1459536. [CrossRef]
60. Dong, W.; Lei, X.; Liu, Y. The mediating role of patients' trust between web-based health information seeking and patients' uncertainty in China: Cross-sectional web-based survey. *J. Med. Internet Res.* **2022**, *24*, e25275. [CrossRef]
61. Wang, Z.; Zhang, X.; Han, D.; Zhao, Y.; Ma, L.; Hao, F. How the Use of an Online Healthcare Community Affects the Doctor-Patient Relationship: An Empirical Study in China. *Front. Public Health* **2023**, *11*, 1145749. [CrossRef]
62. Yan, J.; Liang, C.; Zhou, P. Physician's Service Quality and Patient's Review Behavior: Managing Online Review to Attract More Patients. *Internet Res.* **2024**. [CrossRef]
63. Wang, Q.; Wang, H.; Wang, S.; Zhang, W. How does the design of consultation pages affect patients' perception of physician authority and willingness to seek offline treatment: A randomized controlled trial. *Behav. Sci.* **2023**, *13*, 584. [CrossRef]
64. Shen, T.; Li, Y.; Chen, X. A systematic review of online medical consultation research. *Healthcare* **2024**, *12*, 1687. [CrossRef] [PubMed]
65. Yang, H.; Guo, X.; Wu, T. Exploring the Influence of the Online Physician Service Delivery Process on Patient Satisfaction. *Decis. Support Syst.* **2015**, *78*, 113–121. [CrossRef]
66. Woolley, F.R.; Kane, R.L.; Hughes, C.C.; Wright, D.D. The effects of doctor-patient communication on satisfaction and outcome of care. *Soc. Sci. Med. A* **1978**, *12*, 123–128. [CrossRef]
67. Woo, S.; Choi, M. Medical service quality, patient satisfaction and intent to revisit: Case study of public hub hospitals in the Republic of Korea. *PLoS ONE* **2021**, *16*, e0252241. [CrossRef]
68. Li, J.; Bao, X.; Liu, X.; Ma, L. The Impact of Joining a Team on the Initial Trust in Online Physicians. *Healthcare* **2020**, *8*, 33. [CrossRef]

69. Wei, W.; Ma, X.; Wang, J. Adaptive Experiments Toward Learning Treatment Effect Heterogeneity. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 2025; *forthcoming*. [CrossRef]
70. Cohen, C.; Pignata, S.; Bezak, E.; Tie, M.; Childs, J. Workplace Interventions to Improve Well-being and Reduce Burnout for Nurses, Physicians and Allied Healthcare Professionals: A Systematic Review. *BMJ Open* **2023**, *13*, e071203. [CrossRef]
71. Forbus, J.J.; Berleant, D. Discrete-Event Simulation in Healthcare Settings: A Review. *Modelling* **2022**, *3*, 417–433. [CrossRef]
72. Chen, Z.; Song, Q.; Wang, A.; Xie, D.; Qi, H. Study on the relationships between doctor characteristics and online consultation volume in the online medical community. *Healthcare* **2022**, *10*, 1551. [CrossRef]
73. Jiang, H.; Mi, Z.; Xu, W. Online medical consultation service–oriented recommendations: Systematic review. *J. Med. Internet Res.* **2024**, *26*, e46073. [CrossRef] [PubMed]
74. Chen, J.; Lan, Y.-C.; Chang, Y.-W.; Chang, P.-Y. Exploring doctors' willingness to provide online counseling services: The roles of motivations and costs. *Int. J. Environ. Res. Public Health* **2020**, *17*, 110. [CrossRef]
75. Kammrath Betancor, P.; Boehringer, D.; Jordan, J.; Lüchtenberg, C.; Lambeck, M.; Ketterer, M.C.; Reinhard, T.; Reich, M. Efficient patient care in the digital age: Impact of online appointment scheduling in a medical practice and a university hospital on the "no-show"-rate. *Front. Digit. Health* **2025**, *7*, 1567397. [CrossRef]
76. Chen, P. Effects of the entropy weight on TOPSIS. *Expert Syst. Appl.* **2021**, *168*, 114186. [CrossRef]
77. Wilkins, A.S. To lag or not to lag?: Re-evaluating the use of lagged dependent variables in regression analysis. *Political Sci. Res. Methods* **2018**, *6*, 393–411. [CrossRef]
78. O'Malley, A.J. Instrumental variable specifications and assumptions for longitudinal analysis of mental health cost offsets. *Health Serv. Outcomes Res. Methodol.* **2012**, *12*, 254–272. [CrossRef]
79. Sun, L.; Abraham, S. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *J. Econom.* **2021**, *225*, 175–199. [CrossRef]
80. Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data*; MIT Press: Cambridge, MA, USA, 2010.
81. Di Stefano, G.; Lo Piccolo, E.; Cicero, L. How job crafting dimensions differentially moderate the translation of work conditions into stress perceptions. *Behav. Sci.* **2025**, *15*, 793. [CrossRef]
82. Knight, C.; Tims, M.; Gawke, J.; Parker, S.K. When do job crafting interventions work? The moderating roles of workload, intervention intensity, and participation. *J. Vocat. Behav.* **2021**, *124*, 103522. [CrossRef]
83. Li, S.; Meng, B.; Wang, Q. The double-edged sword effect of relational crafting on job well-being. *Front. Psychol.* **2022**, *13*, 713737. [CrossRef] [PubMed]
84. Wolski, M.; Howard, L.; Richardson, J. A Trust Framework for Online Research Data Services. *Publications* **2017**, *5*, 14. [CrossRef]
85. Gerlich, M. Exploring Motivators for Trust in the Dichotomy of Human—AI Trust Dynamics. *Soc. Sci.* **2024**, *13*, 251. [CrossRef]
86. Yan, Y.; Lu, B.; Xu, T. Exploring the Relationship between Host Self-Description and Consumer Purchase Behavior Using a Self-Presentation Strategy. *Systems* **2023**, *11*, 430. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Impacts of Brand Spillover Effect on Sourcing and Quality Disclosure of the Platform's Store Brand Under Asymmetric Information

Yang Tong [1], Zexuan Shi [1] and Jicai Li [2,*]

[1] College of Economics and Management, Zhejiang Normal University, Jinhua 321004, China; tongyang_ty@163.com (Y.T.); shizexuan1900@163.com (Z.S.)
[2] College of Xingzhi, Zhejiang Normal University, Jinhua 321100, China
* Correspondence: leejicai@163.com

Abstract

In reaction to evolving consumer preferences, prominent platforms, such as Amazon and JD, have progressively established proprietary store brands. However, the problems related to the sourcing of store brands and the disclosure of their quality information remain uncertain. To fill this gap, this paper utilizes game theory to develop a supply chain consisting of a national brand manufacturer, a third-party manufacturer, and a platform, focusing on the platform's optimal sourcing strategy—determining whether to source its store brand from the national or third-party manufacturer—while also considering its quality disclosure strategy. We then examine how essential elements, specifically the brand spillover effect and the disclosure cost, influence these strategic decisions. Our research reveals that the quality information disclosure of the store brand occurs when the product quality surpasses a predetermined threshold. Additionally, although the elevated disclosure cost consistently diminishes quality disclosure, the impact of the brand spillover effect on quality disclosure is nonlinear. Finally, the platform's sourcing strategy depends greatly on the brand spillover effect and the disclosure cost. Specifically, when the brand spillover effect is relatively large (small), the platform prefers to source the store brand from the national (third-party) manufacturer; with a moderate brand spillover effect, a higher (lower) disclosure cost encourages the platform to source from the national (third-party) manufacturer.

Keywords: store brand; brand spillover effect; sourcing strategy; quality disclosure; game theory

1. Introduction

The increasing consumer inclination towards store brands has compelled platforms to vigorously expand their product offerings [1]. Industry titans Amazon and JD illustrate this tendency, having introduced their inaugural store brands in 2005 and 2018, respectively. These projects have developed into strong brand portfolios, including JD's J. ZAO, as well as Amazon's Kindle series. Market performance data demonstrates the exceptional commercial success of this strategy; JD's 2022 financial reports suggest that J. ZAO attained a 60% year-on-year sales increase, with over 100 product categories exceeding an average annual sales growth of 300% (https://www.nbd.com.cn/articles/2023-03-17/2716262.html (accessed on 20 June 2025)).

The swift growth of the market poses intricate strategic challenges for platforms concerning supply chain configuration. Platforms encounter a critical sourcing dilemma: the decision to acquire store brands from national manufacturers or specialized third-party producers. The previous strategy, illustrated by Guangdong Xinbao Electric Appliances Holdings Co., Ltd. (a manufacturer for high-end brands such as Xiaomi, which also produces JD's J.ZAO products), utilizes the brand spillover effect [2,3]. This is because collaboration with these famous national manufacturers can aid platforms in establishing consumer trust in their store brands. Therefore, consumers are more likely to regard the quality of store brands as like that of national brands, thus drawing a broader array of prospective consumers for the store brands. As a result, consumer perception of quality parity between store brands and national brands significantly broadens the potential consumer base when products originate from the same manufacturing sources. Nevertheless, many of JD's store brand categories still come from third-party manufacturers [4]. The underlying motivation for these alternative sourcing strategies is not clear.

At the same time, since consumers cannot directly assess product attributes when shopping online, it becomes difficult for them to ascertain the true value of products, exacerbating uncertainty about product quality [5,6]. Moreover, compared to established national brands that have been on the market for some time, consumer uncertainty about the quality of newly introduced store brands is even more pronounced. This uncertainty can deter purchases, negatively impacting businesses. In practice, platforms can disclose product quality information through sales assistance or in-store media. For example, marketing expenses of JD increased by 127.6% to RMB27.0 billion for the second quarter of 2025 from RMB11.9 billion for the second quarter of 2024 (https://ir.jd.com/news-releases/news-release-details/jdcom-announces-second-quarter-and-interim-2025-results (accessed on 18 September 2025)).

Recent studies have examined the operational and marketing management of store brands, the sourcing of store brands, and quality information disclosure. However, these studies primarily focus on symmetric information settings or neglect the interplay between quality information disclosure and sourcing strategies. Although disclosing high-quality information can build the reputation of store brands and help consumers make informed purchasing decisions based on their preferences and needs, it may also intensify competition between store brands and national brands, adversely affecting the platform's profitability. Additionally, compared to sourcing store brands from third-party manufacturers, procuring from well-known manufacturers generates a brand spillover effect, which helps consumers form higher quality expectations. Due to these consumer impression differences between sourcing strategies, this study tackles three key research questions: (1) When does the platform have the incentive to disclose quality information of the store brand? (2) Will the platform choose to source the store brand from the national manufacturer or the third-party manufacturer? (3) How do the brand spillover effect and the disclosure cost affect these strategic decisions?

To address these research questions, we develop a three-part supply chain framework in which a platform distributes a national brand, sourced from a national manufacturer, alongside a store brand offering. The platform must make a crucial sourcing decision regarding its store brand: whether to source from the national manufacturer (Strategy n) or a third-party manufacturer (Strategy t). Under Strategy n, the store brand gains advantages from the brand spillover effect resulting from common manufacturing origins. The quality of the national brand is modeled as common knowledge among consumers due to its established market presence, while the quality of the store brand is characterized by ex ante uncertainty as a new market entrant. The platform must make a crucial decision about disclosing the store brand quality information to mitigate consumer uncertainty.

This study characterizes equilibrium outcomes for all members of the supply chain within the combinatorial framework of sourcing and disclosure strategies. We analyze profit comparisons to identify the platform's optimal quality disclosure and sourcing strategies across different market conditions.

This research uses game theory to analyze the decision-making behaviors and their interactions among the platform and the two manufacturers. Game theory, which elucidates the impact of interaction decisions among multiple decision-makers, has been extensively applied in the literature on store brand supply chain management to investigate the decision-making behavior of supply chain partners and is also utilized in this research [1–4]. Through the examination of the interactive dynamics between the national manufacturer, the third-party manufacturer, and the platform utilizing game theory, we can offer theoretical support and decision-making guidance for the platform in selecting sourcing and quality information disclosure strategies for the store brand.

The analytical results provide three key insights that have important theoretical and managerial implications. The brand spillover effect consistently improves platform profitability in the context of national manufacturer sourcing, yet its impact on quality disclosure is non-monotonic. A stronger brand spillover effect may paradoxically inhibit quality disclosure relative to third-party sourcing. Secondly, the disclosure cost exhibits dual effects: while it generally impedes quality information disclosure across sourcing strategies, it can paradoxically increase the platform's expected profit under certain market conditions. Finally, the platform's optimal sourcing strategy systematically depends on critical parameters: a higher disclosure cost combined with a significant brand spillover effect favors sourcing from the national manufacturer, whereas their absence systematically shifts the preference toward third-party sourcing.

The subsequent sections of this work are organized as follows. Section 2 examines pertinent literature. Section 3 establishes the analytical framework. Section 4 subsequently deduces equilibrium results across different sourcing strategies. Section 5 delineates main findings and managerial insights, while Section 6 offers concluding observations. Appendix A contains mathematical proofs.

2. Literature Review

This study is primarily related to three streams of literature: store brands, the sourcing strategy of store brands, and information disclosure. This section reviews key contributions in these areas and highlights the novelty of this study.

2.1. Store Brands

This research adds to the existing body of work on store brand operational and marketing strategies [7,8]. Previous studies indicate that the retailer can attain Pareto improvements via the introduction of store brands, as evidenced by Ru et al. [9], who found that the retailer in a category fosters mutually beneficial outcomes for both the manufacturer and the retailer. Ru et al. [10] demonstrate that bargaining power significantly influences the quality positioning of store brands, revealing that the stronger retailer tends to offer higher-quality store brands, whereas the weaker retailer chooses lower-quality options. Implications for profitability are further complicated in multi-channel environments, as demonstrated by Xiao et al. [11], where the retailer's profitability is critically dependent on the calibration of store brand quality. Long et al. [12] identify significant welfare trade-offs: although the platform strategically lowers the price of the store brand, their prominent shelf placement results in measurable consumer welfare losses. Chen et al. [13] analyze the effect of implementing the store brand within the live streaming channel and indicate that the store brand typically erodes the manufacturer's profitability. Zhang et al. [14] further

study the strategic interaction between the store and the national brands, and they find that the manufacturer would dynamically adjust the national brand's quality according to investment efficiency to cope with potential competition from the store brand.

However, unlike the above studies that focus on the symmetric information setting, a variety of studies have explored the impact of introducing store brands in scenarios characterized by information asymmetry, such as asymmetries in production cost [15–17], market demand [18], product quality [19], or the consumer acceptance of the store brand [4]. For example, Cao et al. [15] illustrate that the introduction of the store brand in the context of cost information asymmetry leads to Pareto-dominated outcomes, resulting in decreased profitability for both the retailer and the manufacturer. Cao et al. [16] study the optimal contract design strategy under cost information asymmetry. They find that when the national brand is more competitive (i.e., of higher quality or lower cost), the profits of both the national brand manufacturer and the retailer may increase. Luo et al. [17] indicate that the retailer often retains an informational advantage and is hesitant to reveal the actual pricing data of its store brand to the manufacturer. This informational advantage allows the retailer to secure a more advantageous position within the supply chain, resulting in increased earnings. Additionally, Shi and Geng [18] establish threshold conditions for the store brand adoption under different regimes of market uncertainty disclosure. Huang et al. [19] identify factors that influence quality information sharing in a dual-channel environment where the retailer offers both the national and the store brands. Tong and Xiao [4] reveal that the existence of asymmetric consumer acceptance of the store brand has the potential to hinder the platform's ability to acquire the store brand from the national manufacturer.

This research distinguishes itself by making three unique contributions, in contrast to existing literature that primarily focuses on motivations and impacts of the store brand introduction within symmetric information environments (e.g., [9,11]). First, this study shifts the analytical focus to the operational decisions of e-commerce platforms in the context of asymmetric information. Second, a unified framework is developed to examine the selection of concurrent procurement strategies (national versus third-party sourcing) and endogenous quality disclosure. Finally, the brand spillover effect is introduced as a key driver of platform decision-making, highlighting its interaction with the disclosure cost as a crucial factor in optimal strategy formulation.

2.2. Sourcing Strategy of Store Brands

The second pertinent literature stream investigates the store brand sourcing strategy [20–22]. Platforms exhibit distributional advantages in warehousing and logistics; however, they fundamentally rely on manufacturing partners for production capabilities. Recent scholarly research has concentrated on the optimal sourcing configuration of the retailer, highlighting several key determinants: consumer service sensitivity coefficients [20], supply chain pricing power [23], and the intensity of manufacturer competition [24]. For example, Li et al. [20] point out that when consumers are highly sensitive to service, the e-retailer tends to produce store brands themselves. Conversely, when consumers are less sensitive to service, the e-retailer tends to purchase store brands from the third-party manufacturer or the national brand manufacturer. Liao et al. [23] examine the optimization of store brand quality decisions by the retailer, considering the influence of sourcing arrangement and channel power dynamics. Cheng et al. [24] illustrate that the introduction of store brands serves two strategic functions: it not only generates direct revenue but also acts as a bargaining tool to obtain advantageous wholesale conditions from the national manufacturer. This viewpoint is supported by Zheng et al. [25], who demonstrate that although

store brands generally increase the retailer's bargaining power, effective implementation necessitates strategic adjustment to market structures and cost configurations.

Moreover, there are some studies that also explore the sourcing strategy of the retailer for store brands between the national brand manufacturer and the third-party manufacturer. For instance, Li et al. [20] indicate that the optimal purchasing strategy of an e-tailer mainly depends on the production cost of the store brand and the sensitivity of consumers. When the production cost is low, the e-tailer tends to produce the store brand internally; otherwise, when the production cost is high, the e-tailer is more inclined to purchase the store brand from the third-party manufacturer. Cui et al. [26] pointed out that when the brand spillover effect is significant and the market size of the store brand is small, the platform tends to choose to purchase the store brand from the national brand manufacturers rather than the third-party one. This is because the brand spillover effect can significantly enhance the market acceptance of the store brand, thereby expanding its market share.

Different from the above studies, based on the quality information asymmetry between consumers and supply chain members, this paper jointly studies the optimal quality disclosure decision of the platform under different sourcing strategies, and finally obtains its optimal sourcing strategy. We find that the brand spillover effect of sourcing the store brand from the national manufacturer will affect the platform's quality disclosure strategy and then affect its expected profit and preference for different sourcing strategies.

2.3. Information Disclosure

This study also corresponds with the literature investigating quality disclosure tactics in consumer uncertainty. Existing research indicates that while information disclosed by sellers can help consumers identify quality preferences and reduce uncertainty, whether sellers benefit from such disclosure remains unclear. For instance, Guan and Chen [27] study the interplay between the manufacturer's information acquisition and quality disclosure strategies. Yu et al. [28] show that consumers can leverage the disclosed quality information to make more informed purchasing decisions, choosing products that meet their expectations and needs. Kuksov and Lin's [29] seminal work established optimal frameworks for information disclosure when consumers encounter dual ambiguities concerning product features and preference alignment. Ghosh and Galbreth [30] illustrate how customer attention limitations and search expenses affect disclosure decisions, whilst Gu and Xie [31] investigate competitive disclosure motivations that enhance consumer-product alignment. Du et al. [32] explore how firms can strategically disclose quality and price information about new services to shape consumer beliefs and purchasing decisions. Cao et al. [33] compare the different quality disclosure strategies and conclude the manufacturer's incentive of quality disclosure.

In addition, with the strong rise of e-commerce platforms and the fact that e-commerce platforms can collect a large amount of consumer information, many researchers have also begun to focus on information disclosure or sharing on e-commerce platforms. For example, Jiang et al. [34] indicate that both suppliers and retailers tend to shift the responsibility for information omissions onto each other to take advantage of the other's costs to enhance consumers' expectations of product quality. Li et al. [35] consider whether the online retailer can benefit from disclosing product demand when the manufacturer not only provides products but also competes with them. Based on the competition in the upstream of the supply chain, Liu et al. [36] study the optimal information sharing strategy of the e-commerce platform when multiple sellers simultaneously sell products through a single e-commerce platform. Tsunoda and Zennyo [37] investigate the relationship between the information sharing strategy of the e-commerce platform and the choice of cooperation mode of the manufacturer and find that information sharing can promote the manufacturer

to choose the agency mode. Similarly, Zha et al. [38] show that the information sharing strategy can promote the manufacturer to cooperate with the e-commerce platform through a dual-channel mode. Ha et al. [39] examine the relationship between e-commerce platform information sharing strategy and manufacturer channel strategy.

However, the above studies do not consider the impact of different sourcing strategies adopted by the e-commerce platform on its quality disclosure strategies. When sourcing the store brand from the national manufacturer can help enhance brand awareness, the brand spillover effect will have a significant impact on the quality disclosure strategy of the platform through its pricing strategy, and ultimately affect its sourcing strategy by affecting the expected profit.

2.4. The Contribution of This Study

In summary, this study aims to explore how the platform decides on quality disclosure and sourcing strategies for its store brand in the presence of asymmetric information. Although numerous studies have examined the operational and marketing management of store brands (e.g., [9,11]), the sourcing of store brands (e.g., [23,24]), and quality information disclosure (e.g., [34]), they primarily focus on symmetric information settings or neglect the interplay between quality information disclosure and sourcing strategies. In contrast, our study extends this literature by utilizing a game-theoretic model that incorporates the brand spillover effect and asymmetric quality information to analyze the optimal quality information disclosure and sourcing strategies for the platform. We show that, although sourcing store brands from the national manufacturer can enhance consumer quality expectations through the brand spillover effect, this positive effect may be offset by the cost and strategic implications of quality disclosure. As a result, the brand spillover effect may inadvertently complicate the platform's decision to disclose quality information, thus influencing its sourcing strategy. This insight highlights the interplay between the brand spillover effect and quality disclosure strategy, offering a new perspective on how the platform manages information disclosure and sourcing strategies. Therefore, our research not only contributes to the existing literature on the store brand sourcing strategy and information disclosure but also provides valuable insights for platforms to design effective information disclosure and sourcing strategies of store brands. Table 1 summarizes the differences between previous studies and our research.

Table 1. The gap between our research and previous studies.

	Store Brand	**Asymmetric Information**	**Sourcing Strategy of Store Brand**	**Information Disclosure**	**Brand Spillover**
Chen et al. [13]	√				
Zhang et al. [14]	√				
Cao et al. [16]	√	√		√	
Luo et al. [17]	√	√		√	
Cheng et al. [24]	√		√		
Zheng et al. [25]	√		√		
Cui et al. [27]	√		√		√
Du et al. [32]				√	√
Cao et al. [33]	√		√		√
This paper	√	√	√	√	√

3. Model Description and Assumptions

We model a tripartite supply chain comprising a national brand manufacturer (M), a third-party manufacturer (T), and an online marketplace platform (P). Following Cao

et al. [16] and Xiao et al. [11], we suppose that the national brand manufacturer M produces the national brand (N) distributed through the platform. To capture additional market segments, the platform can introduce an alternative store brand (S) that competes with the national brand. The platform faces a critical sourcing decision for the store brand: sourcing either from manufacturer M (Strategy n) or from manufacturer T (Strategy t). For analytical tractability and without loss of generality, we normalize production costs for both brands to zero.

Assume the two brands engage in price competition, with their demand functions as follows:

$$D_N = q_N - p_N + bp_S$$

$$D_S = q_S + \alpha(q_N - q_S) - p_S + bp_N$$

Let $b \in [0, 1)$ denote the degree of substitutability between the national brand and the store brand. D_N and D_S (p_N and p_S) denote the demands (retail prices) for the two brands, respectively. q_N and q_S represent their respective quality levels. Without loss of generality, we assume $q_N > q_S$, and $q_N - q_S$ indicates the quality difference between the two brands. When the platform sources the store brand from the national manufacturer (Strategy n), consumers' perceived quality exhibits a spillover premium: $q_S + \alpha(q_N - q_S)$, where $\alpha \in [0, 1)$ quantifies the brand spillover effect [2–4]. This formulation captures two essential properties: (1) the national brand's reputation confers a perceived quality enhancement on the store brand; (2) the store brand's perceived quality remains strictly dominated regardless of the brand spillover effect. The channel structure is shown in Figure 1.

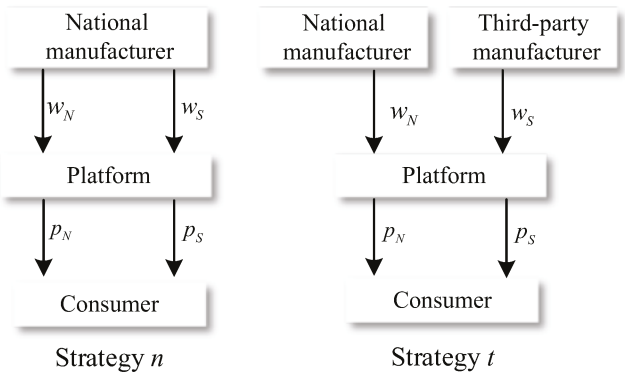


Figure 1. Channel structures under different sourcing strategies.

Given the national brand's established market presence, its quality level constitutes common knowledge among consumers and is normalized to be 1, i.e., $q_N = 1$. Conversely, the store brand's quality—modeled as a random variable $q_S \sim U[0, 1]$—remains unobservable to consumers upon introduction. The platform, possessing private information about q_S prior to product launch, strategically decides whether to disclose this quality information. Let $\widetilde{q}_S$ denote consumers' posterior quality perception, where $\widetilde{q}_S = q_S$ under disclosure and $\widetilde{q}_S = \overline{q}_S$ under non-disclosure. Consistent with rational expectations literature [34], the platform will disclose only when quality exceeds an endogenous threshold $\hat{q}_S$ where marginal disclosure benefits outweigh costs. Thus, if the platform does not disclose quality information, we have $\overline{q}_S = E[q_S | q_S \leq \hat{q}_S] = \hat{q}_S/2$. Additionally, information disclosure must be truthful—either the true quality is disclosed, or the platform remains silent [40,41]. Let $d_R \in \{0, 1\}$ represents the platform's disclosure strategy, where $d_R = 1$ denotes disclosure and $d_R = 0$ denotes non-disclosure. If the platform chooses to disclose, it incurs a fixed disclosure cost c. To ensure the platform has an incentive to disclose in

equilibrium, assume $c \leq \hat{c} = \frac{(1-\alpha)(3+\alpha+4b)}{64(1-b^2)}$; otherwise, a sufficiently large disclosure cost will always inhibit the platform from disclosing quality information.

The sequence of the game is illustrated in Figure 2. In Stage 1, following Li et al. [20] and Cui et al. [26], we assume that before observing the store brand's quality information, the platform decides its sourcing strategy (i.e., Strategy n or Strategy t). In Stage 2, the platform decides whether to disclose the store brand's quality information. In Stage 3, after observing the quality information, the platform and manufacturers determine optimal prices. Specifically, under Strategy n, the national manufacturer first sets wholesale prices w_N and w_S for both brands, followed by the platform sets retail prices p_N and p_S. Under Strategy t, the national manufacturer and the third-party manufacturer simultaneously set wholesale prices w_N and w_S, respectively, while the platform sets retail prices p_N and p_S.

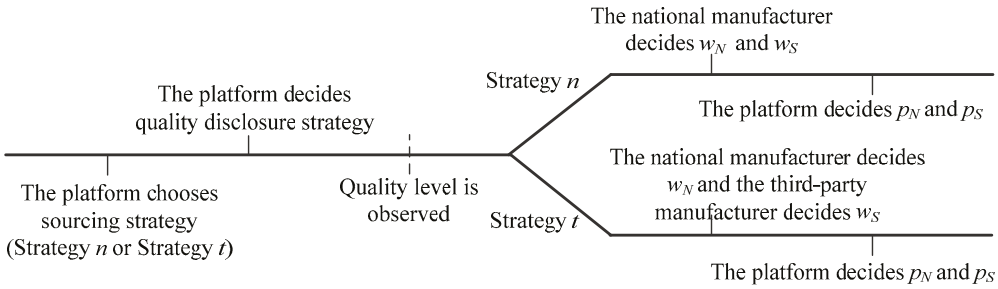


Figure 2. Game sequence.

Notably, this model assumes the platform decides its sourcing strategy before observing the quality level for the store brand. This is because sourcing is a relatively long-term decision; it can be changed occasionally, but manufacturers generally like contracts that span a year or longer. This is because they must allocate resources to meet the demands of producing the store brand. This assumption aligns with the literature on store brand sourcing [4,16,19]. Table 2 summarizes the notations employed in this paper.

Table 2. Summary of notations.

Notation	Explanation
Decision variable	
w_N	The wholesale price of the national brand
w_S	The wholesale price of the store brand
p_N	The retail price of the national brand
p_S	The retail price of the store brand
Parameters	
$D_{N/S}$	The demand for the two brands
$q_{N/S}$	The quality levels for the two brands
b	The degree of substitutability between the two brands, $b \in [0,1)$
α	The degree of the brand spillover effect, $\alpha \in [0,1)$
$\tilde{q}_S$	The consumers' posterior quality perception under disclosure strategy
$\bar{q}_S$	The consumers' posterior quality perception under non-disclosure strategy
$\hat{q}_S$	The threshold where marginal disclosure benefits outweigh the cost
d_R	The platform's strategy collection (Disclosure/Non-disclosure), $d_R \in \{0,1\}$
c	The fixed disclosure cost

4. Model Analysis

4.1. Strategy t

Under Strategy t, the platform sources the store brand from manufacturer T, and thus, $\alpha = 0$. Given the quality disclosure strategy, i.e., $q_S = \widetilde{q}_S$, the profit functions for the manufacturers and the platform are as follows:

$$\pi_P = (p_N - w_N)(1 - p_N + bp_S) + (p_S - w_S)\left(\widetilde{q}_S - p_S + bp_N\right) - d_R c$$

$$\pi_M = w_N(1 - p_N + bp_S)$$

$$\pi_T = w_S\left(\widetilde{q}_S - p_S + bp_N\right)$$

By solving the firms' problems, we derive Lemma 1.

Lemma 1. *Under Strategy t, given $\widetilde{q}_S$, the optimal wholesale prices are $w_N^{t*} = \frac{2+b\widetilde{q}_S}{4-b^2}$ and $w_S^{t*} = \frac{b+2\widetilde{q}_S}{4-b^2}$, and the optimal retail prices are $p_N^{t*} = \frac{6-3b^2+b\widetilde{q}_S\left(5-2b^2\right)}{8-10b^2+2b^4}$ and $p_S^{t*} = \frac{b\left(5-2b^2\right)+3\widetilde{q}_S\left(2-3b^2\right)}{8-10b^2+2b^4}$. The equilibrium sales volumes are $D_N^{t*} = \frac{2+b\widetilde{q}_S}{8-2b^2}$ and $D_S^{t*} = \frac{b+2\widetilde{q}_S}{8-2b^2}$, respectively.*

Lemma 1 establishes that an enhancement in consumers' perceived quality of the store brand (i.e., a larger $\widetilde{q}_S$) induces strictly increasing wholesale price responses from both manufacturers. Consequently, the platform optimally adjusts retail prices upward, demonstrating monotonicity in $\widetilde{q}_S$. Crucially, while elevated retail prices exert downward pressure on demand, this effect is dominated by the positive demand elasticity with respect to quality perception. Thus, equilibrium sales volumes for both brands maintain their monotonically increasing relationship with $\widetilde{q}_S$.

Proposition 1. *Under Strategy t, there exists $\hat{q}_S^t$ such that:*

(1) The platform will disclose the store brand's quality information if and only if $q_S \geq \hat{q}_S^t$.

(2) $\frac{\partial \hat{q}_S^t}{\partial c} > 0$ and $\frac{\partial \hat{q}_S^t}{\partial b} < 0$ always hold.

Proposition 1(1) characterizes the platform's optimal disclosure strategy under Strategy t. Building on Lemma 1's established monotonic relationship between platform profit and perceived quality, the platform adopts a threshold disclosure rule: quality information is revealed if and only if $q_S \geq \hat{q}_S^t$. This equilibrium strategy reflects rational economic calculus-disclosure of subthreshold quality levels would both depress demand and incur unnecessary disclosure costs c, while supra-threshold disclosure generates net positive returns.

Proposition 1(2) identifies two critical determinants of the disclosure threshold $\hat{q}_S^t$: the fixed disclosure cost c and the product substitutability b. The comparative statics reveal: (i) The disclosure threshold increases monotonically with c, as a higher cost erodes the net benefit of quality disclosure, consistent with observed corporate non-disclosure practices; (ii) The threshold decreases with b, since greater substitutability attenuates inter-brand competition, thereby amplifying the marginal profit from quality disclosure.

4.2. Strategy n

Under Strategy n, the platform sources the store brand from manufacturer N, and thus, $\alpha \in [0, 1)$. Again, given the disclosure strategy, i.e., $q_S = \widetilde{q}_S$, the profit functions are:

$$\pi_P = (p_N - w_N)(1 - p_N + bp_S) + (p_S - w_S)\left[\widetilde{q}_S + \alpha\left(1 - \widetilde{q}_S\right) - p_S + bp_N\right] - d_R c$$

$$\pi_M = w_N(1 - p_N + bp_S) + w_S\left[\widetilde{q}_S + \alpha\left(1 - \widetilde{q}_S\right) - p_S + bp_N\right]$$

By solving the firms' problems, we derive Lemma 1.

Lemma 2. *Under Strategy n, given* $\widetilde{q}_S$*, the optimal wholesale prices are* $w_N^{n*} = \frac{1+\alpha b+b\widetilde{q}_S(1-\alpha)}{2-2b^2}$ *and* $w_S^{n*} = \frac{\alpha+b+\widetilde{q}_S(1-\alpha)}{2-2b^2}$*, and the optimal retail prices are* $p_N^{n*} = \frac{3\left[1+\alpha b+b\widetilde{q}_S(1-\alpha)\right]}{4-4b^2}$ *and* $p_S^{n*} = \frac{3\left[\alpha+b+\widetilde{q}_S(1-\alpha)\right]}{4-4b^2}$*. The equilibrium sales volumes are* $D_N^{n*} = \frac{1}{4}$ *and* $D_S^{n*} = \frac{b+2\widetilde{q}_S}{8-2b^2}$ *, respectively.*

Lemma 2 shows that under Strategy n, the impact of $\widetilde{q}_S$ on prices aligns with Strategy t, i.e., w_N^{n*}, w_S^{n*}, p_N^{n*}, and p_S^{n*} are monotonically increasing in $\widetilde{q}_S$. Additionally, as the brand spillover effect strengthens (i.e., a larger α), manufacturer M can charge higher wholesale prices, leading the platform to set higher retail prices as well. Again, the joint positive effects of $\widetilde{q}_S$ and α will dominate the negative effect of the increased retail prices, thus ensuring demand remains non-negatively affected (where D_N^{n*} is independent of $\widetilde{q}_S$ and α, while D_S^{n*} is monotonically increasing in both).

Proposition 2. *Under Strategyn, there exists* $\hat{q}_S^n$ *such that:*

(1) *The platform will disclose the store brand's quality information if and only if* $q_S \geq \hat{q}_S^n$.

(2) $\frac{\partial \hat{q}_S^n}{\partial c} > 0$ *and* $\frac{\partial \hat{q}_S^n}{\partial b} < 0$ *always hold.*

(3) $\frac{\partial \hat{q}_S^n}{\partial \alpha} < 0$ *if* $\alpha \in \left[0, \frac{1-b}{2}\right)$ *and* $c \in \left(0, \frac{1-b-2\alpha}{48-48b}\right)$*; otherwise, we have* $\frac{\partial \hat{q}_S^n}{\partial \alpha} > 0$.

Similarly, Proposition 2 shows that under Strategy n, the platform will only disclose quality information if the store brand's quality exceeds a threshold (i.e., $q_S \geq \hat{q}_S^n$). Otherwise, disclosing low-quality information would harm demand and incur the disclosure cost. Additionally, a higher disclosure cost discourages disclosure, while a greater substitutability encourages it.

Interestingly, Proposition 2(3) reveals that while a stronger spillover effect boosts the platform's profit, it does not always incentivize disclosure. Specifically, when both the brand spillover effect and the disclosure cost are below certain thresholds (i.e., $\alpha < \frac{1-b}{2}$ and $c < \frac{1-b-2\alpha}{48-48b}$), an increase in α encourages disclosure (i.e., $\frac{\partial \hat{q}_S^n}{\partial \alpha} < 0$). Conversely, beyond these thresholds, a higher α discourages disclosure. This is because while the brand spillover effect elevates the store brand's perceived quality, it narrows the quality gap between disclosure and non-disclosure scenarios. When α approaches 1, the quality levels under both strategies converge, making disclosure less beneficial relative to its cost. Thus, under certain conditions, $\hat{q}_S^n$ is monotonically increasing in α.

5. Results and Managerial Insights

Based on the pricing and quality disclosure decisions of the platform under the two sourcing strategies, Section 5.1 firstly analyzes the differences in quality disclosure strategies between the two sourcing strategies. Section 5.2 then examines the platform's optimal sourcing strategy and the influence of parameters such as the brand spillover effect and the fixed disclosure cost on this strategy.

5.1. Quality Disclosure Strategies Comparison

By comparing the quality disclosure thresholds as presented in Propositions 1 and 2, we can identify the differences in the platform's disclosure strategies under the two sourcing strategies, as well as whether parameters like the brand spillover effect and the disclosure cost influence these differences.

Proposition 3. *Comparing the disclosure thresholds under Strategy t and Strategy n, we find that when* $b \in [0, 0.167)$, $\alpha \in (\alpha_1, \alpha_2)$, *and* $c \in (0, c_1)$, *we have* $\hat{q}_S^n < \hat{q}_S^t$; *otherwise, we have* $\hat{q}_S^n \geq \hat{q}_S^t$.

Proposition 3 and Figure 3 reveal that, compared to Strategy t, an increase Strategy n, in the brand spillover effect does not always incentivize the platform to disclose the store brand's quality information when sourcing from a national manufacturer. Specifically, when the substitutability between the two brands is low and the disclosure cost is below a certain threshold, either an excessively weak or excessively strong brand spillover effect will discourage quality disclosure under Strategy n, leading to $\hat{q}_S^n \geq \hat{q}_S^t$. Conversely, a moderate brand spillover effect encourages disclosure. Recall from Proposition 2(3), the disclosure threshold under Strategy n first decreases and then increases with α, which explains this result.

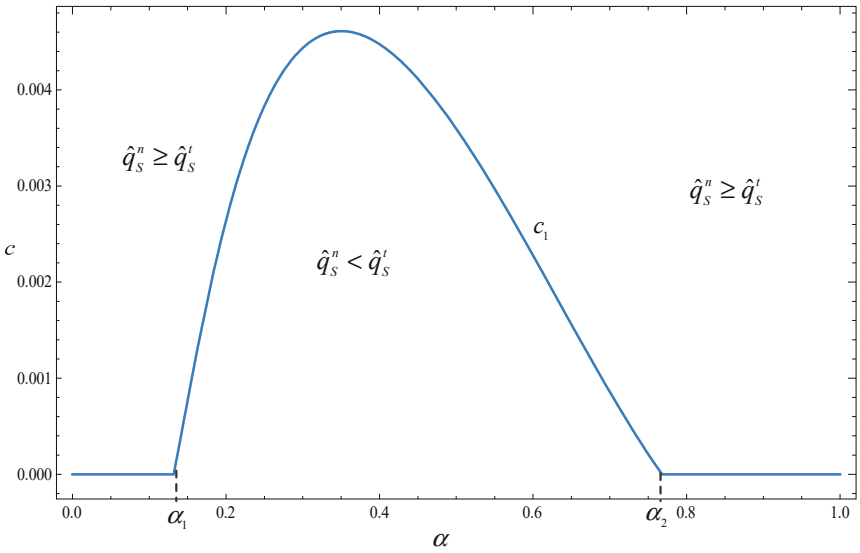


Figure 3. Comparison of disclosure thresholds under Strategy t and where $b = 0.1$.

The findings of Proposition 3 offer important managerial insights for the platform's quality disclosure decision. Although sourcing from a national manufacturer enhances profits through the brand spillover effect, it does not always incentivize quality disclosure. Therefore, when deciding whether to disclose quality information, the platform must consider the interplay of product substitutability, the brand spillover effect, and the disclosure cost.

5.2. The Optimal Sourcing Strategy

Based on the above results, the platform's expected profits under the two sourcing strategies can be calculated as:

$$\pi_P^{t*} = \int_0^{\hat{q}_S^t} \frac{4(4+b^2) + 4b\hat{q}_S^t(8+b^2) + 5b^2(\hat{q}_S^t)^2 + 4(\hat{q}_S^t)^3}{16(1-b^2)(4-b^2)^2} \mathrm{d}q_S + \int_{\hat{q}_S^t}^1 \left[\frac{4+5b^2+2bq_S(8+b^2)+q_S^2(4+5b^2)}{4(1-b^2)(4-b^2)^2} - c \right] \mathrm{d}q_S$$

$$\pi_P^{n*} = \int_0^{\hat{q}_S^n} \frac{4 + [2\alpha + \hat{q}_S^n(1-\alpha)][2\alpha + 4b + \hat{q}_S^n(1-\alpha)]}{64(1-b^2)} \mathrm{d}q_S + \int_{\hat{q}_S^n}^1 \left[\frac{1 + [q_S + \alpha(1-q_S)][2b + q_S + \alpha(1-q_S)]}{16(1-b^2)} - c \right] \mathrm{d}q_S$$

Proposition 4. *Under Strategy n, we have $\frac{\partial \pi_P^{n*}}{\partial \alpha} > 0$ always hold; $\frac{\partial \pi_P^{n*}}{\partial c} < 0$ if $c < c_2$ and $\frac{\partial \pi_P^{n*}}{\partial c} > 0$ otherwise. Under Strategy t, we have $\frac{\partial \pi_P^{t*}}{\partial c} < 0$ if $c < min\{c_3, \hat{c}\}$ and $\frac{\partial \pi_P^{n*}}{\partial c} > 0$ otherwise.*

Intuitively, the platform's expected profit decreases as the disclosure cost rises. However, Proposition 4 shows that once the disclosure cost exceeds a certain threshold, further increases may benefit the platform. This is because a higher disclosure cost reduces the platform's willingness to disclose (i.e., $\frac{\partial \hat{q}_S^t}{\partial c} > 0$ and $\frac{\partial \hat{q}_S^n}{\partial c} > 0$ hold), which mitigates the negative impact of the disclosure cost on the expected profit. Thus, the platform's expected profit may increase with c. Figure 4 illustrates that higher disclosure costs may paradoxically enhance the platform's profitability.

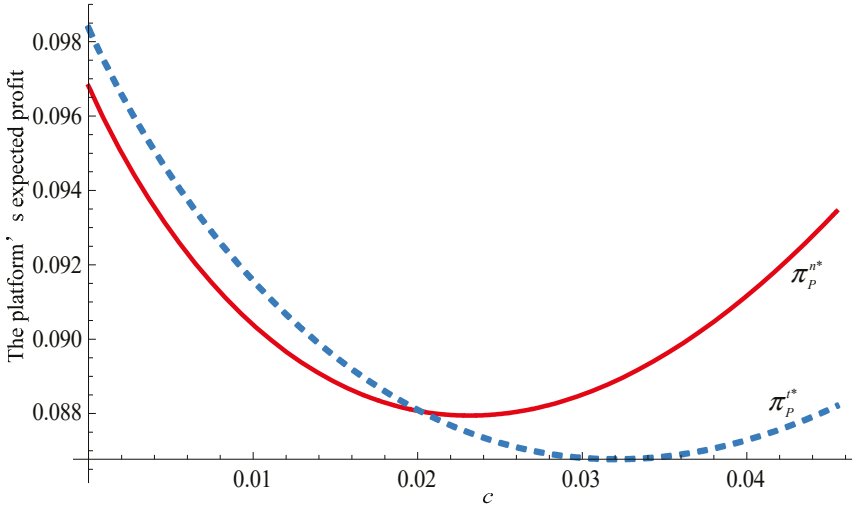


Figure 4. Effect of the disclosure cost c on the platform's expected profit, where $b = 0.1$ and $\alpha = 0.2$.

Proposition 5. *Given $\alpha = 0$, $\pi_P^{t*} > \pi_P^{n*}$ always holds.*

Proposition 5 indicates that in the absence of brand spillover effects, the platform prefers sourcing from a third-party manufacturer. Without the brand spillover effect, Strategy n offers no additional benefits. Moreover, the third-party manufacturer intensifies upstream competition, driving down wholesale prices and benefiting the platform. Thus, when $\alpha = 0$, the platform consistently prefers Strategy t.

Proposition 6. *When the brand spillover effect exists (i.e., $\alpha > 0$), the platform is incentivized to source from a national manufacturer.*

As shown in Figure 5, the platform's sourcing strategy depends on the magnitude of the brand spillover effect and the disclosure cost. Specifically, for the weak spillover effect, the platform prefers third-party procurement, as upstream competition suppresses wholesale prices, boosting profits. In contrast, with the strong brand spillover effect, the platform favors sourcing from a national manufacturer, where spillover-driven profit gains dominate the decision. Interestingly, with the moderate brand spillover effect, the high (low) disclosure cost incentivizes sourcing from a national (third-party) manufacturer. Here, the positive brand spillover effect offsets the disclosure cost, but when this cost exceeds a threshold, the brand spillover advantages become decisive.

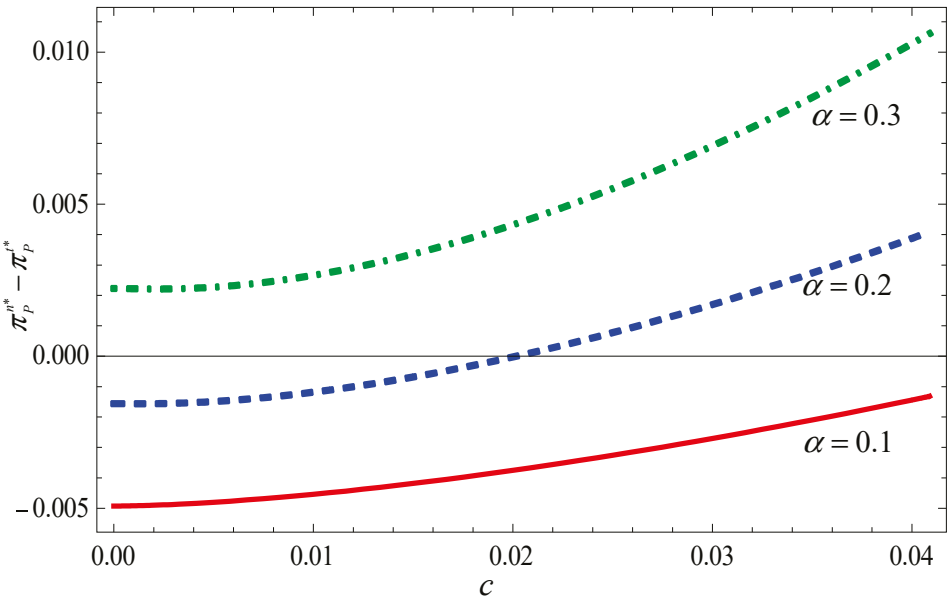


Figure 5. Effect of the disclosure cost on the platform's expected profit difference, where $b = 0.1$.

These findings provide practical guidance for the platform's procurement strategy. While the brand spillover effect enhances profits, the platform must weigh its value against the disclosure cost when selecting a procurement approach.

6. Conclusions

This study investigates the strategic interplay between quality disclosure and sourcing strategies for the platform, introducing store brands to compete with national brands. The analytical framework examines a dual-sourcing scenario where (1) the national brand is exclusively sourced from the national manufacturer, while (2) the store brand can be sourced either from the national manufacturer (enabling the brand spillover effect) or a specialized third-party manufacturer. Key model assumptions reflect fundamental market asymmetries: the national brand's quality is common knowledge due to its market maturity, whereas the store brand's quality remains ex ante unobservable to consumers as a market entrant. The platform strategically determines quality disclosure to resolve this information asymmetry and facilitate consumer decision-making. Through systematic comparison of expected profit functions under alternative strategy combinations, we derive equilibrium solutions for both sourcing and disclosure decisions.

Our analysis highlights several novel findings. First, regarding the strategy of quality information disclosure, our research reveals that the platform has an incentive to disclose this information if the store brand's quality level exceeds a certain threshold; otherwise, disclosure is not profitable. Second, the platform's sourcing strategy is heavily influenced by the brand spillover effect and the disclosure cost. Specifically, when the brand spillover effect is relatively large (small), the platform prefers to source the store brand from the national (third-party) manufacturer. Additionally, with a moderate brand spillover effect, a higher (lower) disclosure cost encourages the platform to source from the national (third-party) manufacturer.

6.1. Theoretical Implications

Our study contributes to the existing literature on quality disclosure and store brand sourcing in several important ways. First, this study extends the research on quality information disclosure strategies for store brands. We show that, compared with Strategy t, an increase in the brand spillover effect does not always motivate the platform to disclose

quality information when sourcing from a national manufacturer. Specifically, when the substitutability between the two brands is low and the disclosure cost is below a certain threshold, either an excessively weak or an excessively strong brand spillover effect discourages quality disclosure under Strategy n. Conversely, a moderate brand spillover effect encourages disclosure.

Second, this study enriches the literature on store brand sourcing decisions. We find that there is no one-size-fits-all approach to store brand sourcing; instead, this decision depends on several key factors. Specifically, when the brand spillover effect is relatively large (small), the platform prefers to source the store brand from the national (third-party) manufacturer. When the brand spillover effect is moderate, a higher (lower) disclosure cost encourages the platform to source from the national (third-party) manufacturer.

6.2. Managerial Implications

These findings provide important managerial insights for retail practitioners concerning quality disclosure and sourcing strategies of the store brand. First, the platform must carefully balance the benefits and drawbacks of disclosing the store brand's quality information. While disclosing high-quality information can enhance the reputation of the store brand and help consumers make more informed purchasing decisions, it may also intensify competition between the store brand and the national brand, thereby reducing the platform's profitability. Moreover, the brand spillover effect resulting from sourcing from the national manufacturer influences consumers' quality expectations, which in turn affects both the platform's expected profit and its preference for quality disclosure. Therefore, when determining the quality disclosure strategy, platforms must weigh the trade-offs between the advantages and disadvantages while also considering their sourcing strategies.

Second, our results shed light on the underlying factors driving the variation in sourcing strategies, namely, why some store brands are sourced from national manufacturers while others are sourced from third-party manufacturers. This critical decision is influenced by parameters such as the brand spillover effect and the disclosure cost. Hence, platforms should carefully take these factors into account when formulating their sourcing strategies.

6.3. Limitations and Suggestions for Future Research

Several promising extensions merit future investigation. First, for analytical tractability, we normalize the production costs of both brands to zero. Future research could extend the model by incorporating differentiated production costs for the two brands. Second, while our analysis focuses on quality uncertainty, introducing demand uncertainty would represent another valuable research direction. Exploring how both types of uncertainty interact could further enhance the understanding of supply chain decisions under imperfect information. Finally, considering quality level as an endogenous decision variable can enrich our study. These extensions would further bridge the gap between theoretical modeling and practical implementation.

Author Contributions: Conceptualization, Y.T. and J.L.; methodology, Y.T. and Z.S.; software, Y.T.; validation, Z.S. and J.L.; formal analysis, Z.S.; investigation, Y.T.; writing—original draft preparation, Z.S.; writing—review and editing, Y.T. and J.L.; project administration, J.L.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Social Science Fund of China (Grant No. 22BGL120).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Proof of Lemma 1. By backward induction, we first solve the platform's retail prices of the national brand and the store brand, respectively. Given $\widetilde{q}_S$, the Hessian matrix of the profit function is $\mathbf{H} = \begin{bmatrix} -2 & 2b \\ 2b & -2 \end{bmatrix}$ is negatively definite because $\mathbf{H}_{11} = -2 < 0$ and $\det(\mathbf{H}) = 4(1-b^2) > 0$ given that $b \in [0,1)$. Hence, π_P is joint concave in p_N and p_S, and we can obtain $p_N(w_N) = \frac{1+b\widetilde{q}_S+w_N(1-b^2)}{2(1-b^2)}$ and $p_S(w_S) = \frac{b+\widetilde{q}_S+w_S(1-b^2)}{2(1-b^2)}$ by solving the first-order optimal condition for p_N and p_S. We then plug $p_N(w_N)$ and $p_S(w_S)$ into the profit functions of the national manufacturer and the third-party manufacturer, we have $\pi_M = \frac{w_N(1-w_N+bw_S)}{2}$ and $\pi_T = \frac{w_S\left(\widetilde{q}_S-w_S+bw_N\right)}{2}$, respectively. As $\frac{\partial^2\pi_M}{\partial w_N^2} = -1 < 0$ and $\frac{\partial^2\pi_T}{\partial w_S^2} = -1 < 0$, π_M (π_T) is concave in w_N (w_S). Therefore, the first-order conditions for w_N and w_S yield $w_N^{t*} = \frac{2+b\widetilde{q}_S}{4-b^2}$ and $w_S^{t*} = \frac{b+2\widetilde{q}_S}{4-b^2}$, respectively. Substituting w_N^{t*} and w_S^{t*} into retail prices, we obtain the equilibrium results that are revealed in Lemma 1. □

Proof of Proposition 1. When the platform discloses quality information of the store brand, we have $\widetilde{q}_S = \hat{q}_S$, leading to the platform's profit with $\frac{\left(4+5b^2\right)\left(1+\hat{q}_S^2\right)+2b\hat{q}_S\left(8+b^2\right)}{4(1-b^2)(4-b^2)^2} - c$; in contrast, when the platform does not disclose quality information, we have $\widetilde{q}_S = \hat{q}_S/2$, leading to the platform's profit with $\frac{\left(4+5b^2\right)\left(4+\hat{q}_S^2\right)+4b\hat{q}_S\left(8+b^2\right)}{16(1-b^2)(4-b^2)^2}$. Therefore, by comparing the platform's profit under different disclosure strategies, we can derive that there exists $\hat{q}_S^t = \frac{2[\sqrt{b^2(8+b^2)^2+12c(4-b^2)^2\left(4+b^2-5b^4\right)}-8b-b^3]}{3(4+5b^2)}$ such that the platform chooses to disclose quality information if and only if $q_S \geq \hat{q}_S^t$; otherwise, the platform will not disclose quality information. By taking the first-order derivatives of $\hat{q}_S^t$ with respect to b and c, we can easily prove that $\frac{\partial\hat{q}_S^t}{\partial c} > 0$ and $\frac{\partial\hat{q}_S^t}{\partial b} < 0$. □

Proof of Lemma 2. Following the same logic as shown in Proof of Lemma 1, we first solve the platform's retail prices of the national brand and the store brand, respectively. Given $\widetilde{q}_S$, the Hessian matrix of the profit function is $\mathbf{H} = \begin{bmatrix} -2 & 2b \\ 2b & -2 \end{bmatrix}$ is negatively definite because $\mathbf{H}_{11} = -2 < 0$ and $\det(\mathbf{H}) = 4(1-b^2) > 0$ given that $b \in [0,1)$. Hence, π_P is joint concave in p_N and p_S, and we can obtain $p_N(w_N) = \frac{1+w_N+b(\widetilde{q}_S-bw_N+\alpha-\widetilde{q}_S\alpha)}{2-2b^2}$ and $p_S(w_S) = \frac{\widetilde{q}_S+b+w_S-b^2w_S+\alpha-\widetilde{q}_S\alpha}{2-2b^2}$ by solving the first-order optimal condition for p_N and p_S. We then plug $p_N(w_N)$ and $p_S(w_S)$ into the profit functions of the national manufacturer, we have $\pi_M = \frac{w_N-{w_N}^2+2bw_Nw_S+w_S\left(\widetilde{q}_S-w_S+\alpha-\widetilde{q}_S\alpha\right)}{2}$. Similarly, Given $\widetilde{q}_S$, the Hessian matrix of the national manufacturer's profit function is $\mathbf{H} = \begin{bmatrix} -1 & b \\ b & -1 \end{bmatrix}$ is negatively definite because $\mathbf{H}_{11} = -1 < 0$ and $\det(\mathbf{H}) = 1-b^2 > 0$ given that $b \in [0,1)$. Therefore, π_M is joint concave in w_N and w_S, and we can obtain $w_N^{n*} = \frac{1+\alpha b+b\widetilde{q}_S+\widetilde{q}_S(1-\alpha)}{2-2b^2}$ and $w_S^{n*} = \frac{\alpha+b-\widetilde{q}_S(1-\alpha)}{2-2b^2}$ by solving the first-order optimal condition for w_N and w_S. Substituting w_N^{n*} and w_S^{n*} into retail prices, we obtain the equilibrium results that are revealed in Lemma 2. □

Proof of Proposition 2. As the same logic as shown in the Proof of Proposition 1, when the platform discloses quality information of the store brand, we have $\widetilde{q}_S = \hat{q}_S$, leading to the platform's profit with $\frac{1+[\hat{q}_S(1-\alpha)+\alpha][2b+\hat{q}_S(1-\alpha)+\alpha]}{16(1-b^2)} - c$; in contrast, when the platform does not disclose quality information, we have $\widetilde{q}_S = \hat{q}_S/2$, leading to the platform's profit with $\frac{4+[\hat{q}_S(1-\alpha)+2\alpha][4b+\hat{q}_S(1-\alpha)+2\alpha]}{64(1-b^2)}$. Then, by comparing the platform's profit under different disclosure strategies, we can derive that there exists $\hat{q}_S^n = \frac{2[\sqrt{(\alpha+b)^2+48c(1-b^2)}-\alpha-b]}{3(1-\alpha)}$ such that the platform chooses to disclose quality information if and only if $q_S \geq \hat{q}_S^n$; otherwise, the platform will not disclose quality information. Then, By taking the first-order derivatives of $\hat{q}_S^n$ with respect to c, b and α, we can easily prove that $\frac{\partial \hat{q}_S^n}{\partial c} > 0$ and $\frac{\partial \hat{q}_S^n}{\partial b} < 0$ always hold while $\frac{\partial \hat{q}_S^n}{\partial \alpha} < 0$ holds if and only if $\alpha \in \left[0, \frac{1-b}{2}\right)$ *and* $c \in \left(0, \frac{1-b-2\alpha}{48-48b}\right)$; otherwise, we have $\frac{\partial \hat{q}_S^n}{\partial \alpha} > 0$. □

Proof of Proposition 3. $\hat{q}_S^n - \hat{q}_S^t = \frac{2}{3}\left[\frac{\sqrt{(\alpha+b)^2+48c(1-b^2)}-\alpha-b}{1-\alpha} - \frac{\sqrt{b^2(8+b^2)^2+12c(4-b^2)^2(4+b^2-5b^4)}-b(8+b^2)}{4+5b^2}\right]$. Then, taking the first-order derivatives of $\hat{q}_S^n - \hat{q}_S^t$ with respect to c, we have

$$\frac{\partial(\hat{q}_S^n - \hat{q}_S^t)}{\partial c} = \frac{4(1-b^2)(4-b^2)^2}{\sqrt{b^2(8+b^2)^2+12c(4-b^2)^2(4+b^2-5b^4)}} - \frac{16(1-b^2)}{(1-\alpha)\sqrt{(\alpha+b)^2+48c(1-b^2)}}$$

Therefore, when $b \in [0.167, 1)$, given that $\alpha \in [0, 1)$ and $c \in (0, \hat{c}]$, we have $\frac{\partial(\hat{q}_S^n - \hat{q}_S^t)}{\partial c} > 0$ always holds. Thus, $\hat{q}_S^n - \hat{q}_S^t \geq \hat{q}_S^n - \hat{q}_S^t\big|_{c=0} = 0$, i.e., $\hat{q}_S^n \geq \hat{q}_S^t$ always holds. When $b \in [0, 0.167)$, $\hat{q}_S^n < \hat{q}_S^t$ holds when $\alpha \in (\alpha_1, \alpha_2)$, and $c \in (0, c_1)$; otherwise, we have $\hat{q}_S^n \geq \hat{q}_S^t$; where $\alpha_1 = \frac{1-b}{2} - \frac{\sqrt{16-96b+8b^2-32b^3-7b^4+2b^5+b^6}}{8-2b^2}$, $\alpha_2 = \frac{1-b}{2} + \frac{\sqrt{16-96b+8b^2-32b^3-7b^4+2b^5+b^6}}{8-2b^2}$, and $c_1 = \frac{(1-\alpha)\left[4b(1-b)-\alpha(2+b)^2\right]\left[b\left(16+12b^2-b^4\right)-\alpha(1-\alpha-b)\left(4+b^2\right)^2\right]}{3(1-b)\left[b^4(1-\alpha)^2-4b^2(7-4\alpha+2\alpha^2)-16\alpha(2-\alpha)\right]^2}$. □

Proof of Proposition 4. The expected profits of the platform under Strategy t and Strategy n are $\pi_P^{t*} = \frac{64+4b(24+20b+3b^2)-(4+5b^2)(\hat{q}_S^t)^3}{48(1-b^2)(4-b^2)^2} - c(1-\hat{q}_S^t)$ and $\pi_P^{n*} = \frac{4(4+\alpha+\alpha^2)+12b(1+\alpha)-(1-\alpha)^2(\hat{q}_S^n)^3}{192(1-b^2)} - c(1-\hat{q}_S^n)$, respectively. Taking the first-order derivatives of π_P^{t*} with respect to c, we have $\frac{\partial \pi_P^{t*}}{\partial c} = 1 - \hat{q}_S^t + \frac{\partial \hat{q}_S^t}{\partial c}\left[c - \frac{(4+5b^2)(\hat{q}_S^t)^2}{16(1-b^2)(4-b^2)^2}\right]$, where $\frac{\partial \hat{q}_S^t}{\partial c} = \frac{4(1-b^2)(4-b^2)^2}{\sqrt{b^2(8+b^2)^2+12c(4-b^2)^2(4+b^2-5b^4)}}$. Therefore, $\frac{\partial \pi_P^{t*}}{\partial c} < 0$ if $c < \min\{c_3, \hat{c}\}$ and $\frac{\partial \pi_P^{n*}}{\partial c} > 0$ otherwise. Similarly, taking the first-order derivatives of π_P^{n*} with respect to α and c, we have $\frac{\partial \pi_P^{n*}}{\partial c} = 1 - \hat{q}_S^n + \frac{\partial \hat{q}_S^n}{\partial c}\left[c - \frac{3(1-\alpha)^2(\hat{q}_S^n)^2}{192(1-b^2)}\right]$ and $\frac{\partial \pi_P^{n*}}{\partial \alpha} = \frac{2+\alpha+6b}{96(1-b^2)} + \frac{2(1-\alpha)(\hat{q}_S^n)^3}{192(1-b^2)} - \frac{\partial \hat{q}_S^n}{\partial \alpha}\left[\frac{3(1-\alpha)^2(\hat{q}_S^n)^2}{192(1-b^2)} - c\right]$, where $\frac{\partial \hat{q}_S^n}{\partial c} = \frac{16(1-b^2)}{(1-\alpha)\sqrt{(\alpha+b)^2+48c(1-b^2)}}$ and $\frac{\partial \hat{q}_S^n}{\partial \alpha} = \frac{2(1+b)[\alpha+b+48c(1-b)]}{3(1-\alpha)^4\sqrt{(\alpha+b)^2+48c(1-b^2)}} - \frac{2(1-b-2\alpha)}{3(1-\alpha)^2}$. Therefore, $\frac{\partial \pi_P^{n*}}{\partial \alpha} > 0$ always holds while $\frac{\partial \pi_P^{n*}}{\partial c} < 0$ if $c < c_2$. Where $c_2 = \frac{1}{4096}\left[32 - \frac{13(1-\alpha)^2}{1+b} + \frac{35-86\alpha-13\alpha^2}{1-b} + \frac{2(9+4b-5\alpha)\sqrt{9-\alpha(10-17\alpha)+8b(1+3\alpha)+16b^2}}{1-b^2}\right]$ and $c_3 = \frac{(36-32b+45b^2-4b^3)(12+32b+15b^2+4b^3)}{512(1-b^2)(4+5b^2)(4-b^2)^2} + \frac{(36+32b+45b^2+4b^3)\sqrt{144+256b+1384b^2+352b^3+481b^4+40b^5+16b^6}}{512(1-b^2)(4+5b^2)(4-b^2)^2}$. □

Proof of Proposition 5. Given $\alpha = 0$, we have $(\pi_P^{t*} - \pi_P^{n*})|_{\alpha=0} = \frac{b[48+b(4-b)(28+16b+3b^2)]}{48(1-b^2)(4-b^2)^2} - \frac{(4-b^2)^2(\hat{q}_S^n|_{\alpha=0})^3+4(4+5b^2)(\hat{q}_S^t)^3}{192(1-b^2)(4-b^2)^2} - c(\hat{q}_S^n|_{\alpha=0} - \hat{q}_S^t)$. It can be easily verified that $(\pi_P^{t*} - \pi_P^{n*})|_{\alpha=0} > 0$ always holds when $b \in [0, 1)$ and $c \in (0, \hat{c}]$. □

References

1. Ning, Y.; Tong, Y.; Li, J. Information sharing with uncertain consumer preferences for store brands. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 58. [CrossRef]
2. Wu, H.; Li, G.; Zheng, H.; Zhang, X. Contingent channel strategies for combating brand spillover in a co-opetitive supply chain. *Transp. Res. Part E Logist. Transp. Rev.* **2022**, *164*, 102830. [CrossRef]
3. Wu, X.; Zhang, F.; Zhou, Y. Brand spillover as a marketing strategy. *Manag. Sci.* **2022**, *68*, 5348–5363. [CrossRef]
4. Tong, Y.; Xiao, T. National or third-party manufacturer? Sourcing strategy of a dominant platform: Signaling game's perspective. *Omega* **2024**, *124*, 103016. [CrossRef]
5. Hong, Y.; Pavlou, P.A. Product fit uncertainty in online markets: Nature, effects, and antecedents. *Inf. Syst. Res.* **2014**, *25*, 328–344. [CrossRef]
6. Hao, L.; Tan, Y. Who wants consumers to be informed? Facilitating information disclosure in a distribution channel. *Inf. Syst. Res.* **2019**, *30*, 34–49. [CrossRef]
7. Jin, Y.; Wu, X.; Hu, Q. Interaction between channel strategy and store brand decisions. *Eur. J. Oper. Res.* **2017**, *256*, 911–923. [CrossRef]
8. Alan, Y.; Kurtuluş, M.; Wang, C. The role of store brand spillover in a retailer's category management strategy. *Manuf. Serv. Oper. Manag.* **2019**, *21*, 620–635. [CrossRef]
9. Ru, J.; Shi, R.; Zhang, J. Does a store brand always hurt the manufacturer of a competing national brand? *Prod. Oper. Manag.* **2015**, *24*, 272–286. [CrossRef]
10. Ru, J.; Sethi, S.; Shi, R.; Zhang, J. Channel power shift and store brand introduction. *Transp. Res. Part E Logist. Transp. Rev.* **2023**, *179*, 103259. [CrossRef]
11. Xiao, Y.; Niu, W.; Zhang, L.; Xue, W. Store brand introduction in a dual-channel supply chain: The roles of quality differentiation and power structure. *Omega* **2023**, *116*, 102802. [CrossRef]
12. Long, F.; Amaldoss, W. Self-preferencing in e-commerce marketplaces: The role of sponsored advertising and private labels. *Mark. Sci.* **2024**, *43*, 925–952. [CrossRef]
13. Chen, Q.; Yan, X.; Bian, Y.; Han, X. Live streaming channel and product assortment with both national and store brand products. *Omega* **2025**, *131*, 103212. [CrossRef]
14. Zhang, J.; Zhang, J.; Yang, R.; Yue, X. Quality strategy to confront store brand encroachment. *Eur. J. Oper. Res.* **2025**, *328*, 832–844. [CrossRef]
15. Cao, Z.; Wang, Y.; Zhao, J.; Min, J. Store brand introduction and quantity decision under asymmetric cost information in a retailer-led supply chain. *Comput. Ind. Eng.* **2021**, *152*, 106995. [CrossRef]
16. Cao, X.; Fang, X.; Xiao, G.; Yang, N. Optimal contract design for a national brand manufacturer under store brand private information. *Manuf. Serv. Oper. Manag.* **2023**, *25*, 1835–1854. [CrossRef]
17. Luo, Y.; Sun, X.; Yue, X. Store brand entry with asymmetric cost information. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *192*, 103790. [CrossRef]
18. Shi, C.L.; Geng, W. To introduce a store brand or not: Roles of market information in supply chains. *Transp. Res. Part E Logist. Transp. Rev.* **2021**, *150*, 102334. [CrossRef]
19. Huang, S.; Zhang, X.; Guan, X.; Yi, Z. Quality information disclosure with retailer store brand introduction in a supply chain. *Comput. Ind. Eng.* **2022**, *171*, 108475. [CrossRef]
20. Li, H.; Chen, H.; Chai, J.; Shi, V. Private label sourcing for an e-tailer with agency selling and service provision. *Eur. J. Oper. Res.* **2023**, *305*, 114–127. [CrossRef]
21. Assarzadegan, P.; Hejazi, S.R.; Rasti-Barzoki, M. A stackelberg differential game theoretic approach for analyzing coordination strategies in a supply chain with retailer's premium store brand. *Ann. Oper. Res.* **2024**, *336*, 1519–1549. [CrossRef]
22. Zhang, C.; Zhang, Y.; Zhu, Y.; Ren, X.; Sun, S. Sourcing strategy for brand extension with social influence. *Oper. Res.* **2025**, *25*, 76. [CrossRef]
23. Liao, B.; Yano, C.A.; Trivedi, M. Optimizing store-brand quality: Impact of choice of producer and channel price leadership. *Prod. Oper. Manag.* **2020**, *29*, 118–137. [CrossRef]
24. Cheng, R.; Ma, W.; Ke, H. Store-brand introduction and production arrangement in the presence of multiple retailers. *Rairo-Oper. Res.* **2020**, *54*, 827–843. [CrossRef]
25. Zheng, Q.; Jang, H.; Pan, X.A. Store-brand introduction and multilateral contracting. *Manuf. Serv. Oper. Manag.* **2022**, *24*, 467–468. [CrossRef]
26. Cui, H.; Zheng, B.; Li, S.; Xue, J. Interaction and optimal decisions of sourcing strategy and selling mode in online platforms. *Int. Trans. Oper. Res.* **2025**. [CrossRef]
27. Guan, X.; Chen, Y.J. The interplay between information acquisition and quality disclosure. *Prod. Oper. Manag.* **2017**, *26*, 389–408. [CrossRef]

28. Yu, Y.; He, Y.; Guo, X.; Li, D.; Huang, H. Quality disclosure strategy with asymmetric demand information in food supply chains. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *183*, 103427. [CrossRef]
29. Kuksov, D.; Lin, Y. Information provision in a vertically differentiated competitive marketplace. *Mark. Sci.* **2010**, *29*, 122–138. [CrossRef]
30. Ghosh, B.; Galbreth, M.R. The impact of consumer attentiveness and search costs on firm quality disclosure: A competitive analysis. *Manag. Sci.* **2013**, *59*, 2604–2621. [CrossRef]
31. Gu, Z.; Xie, Y. Facilitating fit revelation in the competitive market. *Manag. Sci.* **2013**, *59*, 1196–1212. [CrossRef]
32. Du, S.; Peng, X.; Nie, T.; Zhu, Y. Information disclosure and pricing for services with congestion: Extension vs. Encroachment. *Decis. Support Syst.* **2024**, *176*, 114064. [CrossRef]
33. Cao, H.; Guan, X.; Fan, T.; Zhou, L. The acquisition of quality information in a supply chain with voluntary vs. mandatory disclosure. *Prod. Oper. Manag.* **2020**, *29*, 595–616. [CrossRef]
34. Jiang, Y.; Guan, X.; Bian, Y.; Huang, S. Information disclosure with endogenous channel structure. *Nav. Res. Logist.* **2023**, *70*, 105–120. [CrossRef]
35. Li, G.; Tian, L.; Zheng, H. Information sharing in an online marketplace with co-opetitive sellers. *Prod. Oper. Manag.* **2021**, *30*, 3713–3734. [CrossRef]
36. Liu, Z.; Zhang, D.J.; Zhang, F. Information sharing on retail platforms. *Manuf. Serv. Oper. Manag.* **2021**, *23*, 606–619. [CrossRef]
37. Tsunoda, Y.; Zennyo, Y. Platform information transparency and effects on third—Party suppliers and offline retailers. *Prod. Oper. Manag.* **2021**, *30*, 4219–4235. [CrossRef]
38. Zha, Y.; Li, Q.; Huang, T.; Yu, Y. Strategic information sharing of online platforms as resellers or marketplaces. *Mark. Sci.* **2023**, *42*, 659–678. [CrossRef]
39. Ha, A.Y.; Luo, H.; Shang, W. Supplier encroachment, information sharing, and channel structure in online retail platforms. *Prod. Oper. Manag.* **2022**, *31*, 1235–1251. [CrossRef]
40. Guo, L. Quality disclosure formats in a distribution channel. *Manag. Sci.* **2009**, *55*, 1513–1526. [CrossRef]
41. Guo, L.; Zhao, Y. Voluntary quality disclosure and market interaction. *Mark. Sci.* **2009**, *28*, 488–501. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

applied sciences

MDPI

Article

Advancing Machine Learning-Based Streamflow Prediction Through Event Greedy Selection, Asymmetric Loss Function, and Rainfall Forecasting Uncertainty

Soheyla Tofighi [1,*], Faruk Gurbuz [2], Ricardo Mantilla [3] and Shaoping Xiao [1,*]

[1] Department of Mechanical Engineering, Iowa Technology Institute, University of Iowa, Iowa City, IA 52242, USA
[2] Turkish Water Institute, Istanbul 34696, Türkiye; gurbuzfrk@gmail.com
[3] Department of Civil Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada; ricardo.mantilla@umanitoba.ca
* Correspondence: soheyla-tofighi@uiowa.edu (S.T.); shaoping-xiao@uiowa.edu (S.X.)

Abstract

This paper advances machine learning (ML)-based streamflow prediction by strategically selecting rainfall events, introducing a new loss function, and addressing rainfall forecast uncertainties. Focusing on the Iowa River Basin, we applied the stochastic storm transposition (SST) method to create realistic rainfall events, which were input into a hydrological model to generate corresponding streamflow data for training and testing deterministic and probabilistic ML models. Long short-term memory (LSTM) networks were employed to predict streamflow up to 12 h ahead. An active learning approach was used to identify the most informative rainfall events, reducing data generation effort. Additionally, we introduced a novel asymmetric peak loss function to improve peak streamflow prediction accuracy. Incorporating rainfall forecast uncertainties, our probabilistic LSTM model provided uncertainty quantification for streamflow predictions. Performance evaluation using different metrics improved the accuracy and reliability of our models. These contributions enhance flood forecasting and decision-making while significantly reducing computational time and costs.

Keywords: hydrology; machine learning; greedy selection; asymmetric peak loss function; uncertainty quantification; LSTM; probabilistic predictive model

1. Introduction

Streamflow forecasting is fundamental for effective decision-making in hydrology and water resource management, impacting flood mitigation strategies, watershed conservation, water supply management, hydropower generation, agricultural development, reservoir management, and various aspects of integrated water resource management. Nonetheless, the intricate and ever-changing nature of hydrological systems, often exhibiting nonlinear behaviors and variability, poses a continuous challenge in achieving precise streamflow predictions [1–3]. Various models, both process-based and data-driven, have been developed to tackle this challenge, each with its advantages and limitations [4].

Process-based models, rooted in the principles of hydrology and meteorology, employ differential equations to describe transport mechanisms in the surface and subsurface of a watershed [5–7]. These models have the advantage of modeling complex physical interactions, including water exchanges with the atmosphere and the movement of water in rivers

and lakes [8]. However, they require extensive parameterization, high-quality input data, and significant computational resources, making them complex and sometimes impractical for large-scale or real-time applications. On the other hand, data-driven models capture statistical dependencies between explanatory inputs and response outputs, bypassing the need for explicit governing equations of the system [3,9]. These models are computationally efficient, adaptable to different conditions, and capable of capturing nonlinear relationships that may be oversimplified in process-based approaches. However, they may suffer from overfitting, data quality issues, and a lack of physical interpretability.

Specifically, the advent of machine learning (ML) and artificial intelligence (AI) has revolutionized the field of hydrology, particularly in streamflow forecasting, by offering new possibilities through enhanced computational capabilities and innovative algorithms. Traditional ML techniques, such as support vector machines (SVMs) and linear regression, have been employed in streamflow forecasting, achieving varying degrees of success [10,11]. However, recent breakthroughs in deep learning (DL), particularly the development of long short-term memory (LSTM) networks—a subset of recurrent neural networks (RNNs)—have shown great promise in capturing the temporal dependencies and complex patterns inherent in hydrological data.

Building on these advances, this study employs an LSTM-based architecture for streamflow forecasting. We acknowledge that the gated recurrent unit (GRU) offers a simpler architecture with fewer parameters, and Transformer-based models have shown strong performance in recent sequence modeling applications. However, compared to LSTMs, GRUs are less effective on complex tasks, while Transformers require substantially higher computational resources [12,13].

Many studies have shown that LSTM models often perform better than both traditional process-based and ML models in predicting streamflow [14–22]. For instance, Kratzert et al. [14] successfully employed an LSTM model for daily runoff prediction by incorporating meteorological data as input, demonstrating that these models outperformed process-based models in certain scenarios. Xiang et al. [15] proposed a sequence-to-sequence (seq2seq) model and demonstrated that this model surpasses traditional ML methods like linear regression, lasso regression, and ridge regression in terms of predictive accuracy and various other performance measures.

Hunt et al. [19] employed an LSTM model to predict streamflow at 10 river gauging stations across various climatic regions of the western United States, training it with meteorological and hydrological data. Their findings indicate that employing LSTM for medium-range streamflow forecasting outperforms traditional process-based models. Moreover, research conducted by Arsenault et al. [20] demonstrated that LSTM models exceeded the performance of traditional hydrological models in comprehensive evaluations across over 148 North American basins. Chen et al. [21] conducted a study employing a hybrid approach that integrated the soil water assessment tool (SWAT) with LSTM in ungauged and poorly gauged watersheds. They used SWAT to produce an extended data series of watershed processes, which, along with climatic data, served as inputs for the LSTM model. Their results demonstrate that this integration significantly improved the reliability and accuracy of streamflow simulations, outperforming the default SWAT simulations in these challenging watershed conditions. Moreover, Tursun et al. [22] performed streamflow analyses in 40 catchment areas within the Yellow River Basin, revealing that LSTM-based models hold considerable promise for accurately modeling river flow in arid regions affected by human interventions. These studies further affirm the robust capabilities of LSTM networks in challenging hydrological applications.

Despite all the advancements, significant challenges persist in streamflow forecasting using ML/DL models. Developing a robust model generally requires a substantial

number of data samples. Furthermore, ML models that rely on traditional loss functions struggle to accurately predict peak streamflow events [14], which are usually driven by complex interactions among various factors, such as extreme weather conditions, rapid snow melt, and land use changes. These models often fail to capture the complexities involved, posing significant challenges in hydrological contexts where precise prediction of peak values is critical for effective flood risk mitigation, emergency responses, and water resource management.

Another critical limitation of these models is their inability to account for the uncertainties inherent in hydrological variables (e.g., rainfall) [23]. They offer point estimates of streamflow values that overlook these uncertainties or errors and fail to convey the associated uncertainty in predictions, potentially leading to less reliable streamflow forecasts [24]. This uncertainty is often expressed using prediction intervals or probabilistic distributions. Significant research in hydrology has focused on developing methods to quantify and represent such uncertainties within conventional modeling approaches [25,26]. A comparable level of dedication is needed to adapt these methodologies for DL models like LSTM. A variety of methods exist to assess prediction uncertainty. Most rely on enhancing a deterministic model with an additional uncertainty quantification component. Among these, ensemble-based modeling is particularly common, where outputs from multiple ensemble members are combined—using equal or weighted averaging—to approximate the probability distribution of the predicted variable [27,28]. Other commonly employed approaches encompass pre-processing and post-processing methods, including those grounded in Bayesian techniques and quantile regression methods [29–31].

To address these challenges, our study aims to introduce several innovative approaches to enhance the accuracy and reliability of streamflow predictions. Given the complexity of hydrological processes and the variability of storm events, traditional observation-based datasets may not always provide sufficient coverage for robust model evaluation. Therefore, after model calibration, we used datasets generated by process-based models instead of direct observations as they efficiently cover a broader range of storm event scenarios compared to USGS observations. This approach ensured that our models were tested under diverse conditions, including extreme events that might be underrepresented in observational records. Additionally, process-based models allowed for a more controlled and systematic demonstration of the proposed approaches, facilitating a comprehensive assessment of their performance and generalizability.

Firstly, to identify the most informative rainfall events and reduce the computational cost and time associated with hydrological simulations for data generation, we employed an active learning (AL) technique. While our previous study [32] preliminarily investigated the effect of training set size on ML model performance, it relied on randomly selecting storm events to generate training sets of varying sizes. The AL-based selection strategy used in this study advances our previous work, demonstrating a clear advantage over the random selection of rainfall events in our experiments.

Second, we proposed a novel asymmetric loss function designed to capture peak streamflow events more accurately. This approach improved the model's ability to predict high-flow values, addressing a critical aspect of hydrological forecasting where symmetric loss functions might fall short. Compared to the pinball loss (PL) function [30], an available asymmetric loss function, the developed loss function yielded better performance from the ML predictive model.

Finally, we developed a probabilistic LSTM (PLSTM) model that integrates rainfall forecast uncertainties, providing a range of probable streamflow outcomes rather than a single deterministic output. Differing from other studies that used ensemble methods [27,28], Bayesian techniques [29,31], and quantile regression [30], our PLSTM probabilistic model-

ing technique uses negative log-likelihood (NLL) to quantify uncertainty directly during model training. Using this technique, the model estimates prediction uncertainty without the need for ensemble methods or Bayesian techniques, offering an efficient yet reliable solution for streamflow predictions. Unlike most existing studies, which assume deterministic and accurate rainfall forecasts, this approach advances by incorporating rainfall forecast errors into the training data, capturing both the expected value and the associated uncertainty of streamflow predictions. This methodology aligns with and extends the current state of the art by applying probabilistic techniques within DL frameworks to address uncertainty in hydrological modeling.

Through these approaches, our goal is to redefine the benchmarks for streamflow forecasting by integrating advanced ML methods with deep hydrological understanding, thereby enhancing the precision and practical applicability of predictions for flood early warning systems, reservoir operation, and other strategic planning.

The remainder of this paper is organized as follows: Section 2 presents a description of the study area, outlines the process-based modeling framework, and details the calibration procedure. Section 3 describes the data generation, including rainfall event selection and dataset creation. Section 4 introduces the LSTM-based encoder–decoder model for deterministic streamflow prediction and demonstrates the advantage of greedy rainfall event selection. Section 5 presents a novel loss function designed to emphasize peak streamflow prediction. Section 6 incorporates rainfall forecasting errors into the training data and develops a probabilistic LSTM model for quantifying uncertainty in streamflow predictions. Finally, Section 7 concludes the paper.

2. Hydrological Model

2.1. Iowa River Basin

The focus area of this study encompasses the Iowa River Basin and its internal river network. Located in the central and eastern regions of Iowa, the Iowa River Basin drains an area of roughly 11,140 km^2 before reaching Columbus Junction. The basin is continuously monitored by more than 20 USGS gauging stations. Over the last twenty years, this region has experienced a series of recurrent flood events, resulting in significant damage to individuals, agricultural activities, and infrastructure across different scales. In 2008, a substantial flood event impacted 86 counties in Iowa, leading to its designation as a federal disaster zone. This event incurred costs totaling USD 11 billion in damages and affected over 41,000 individuals [33]. Figure 1 depicts the Iowa River Basin along with the locations of the USGS stations.

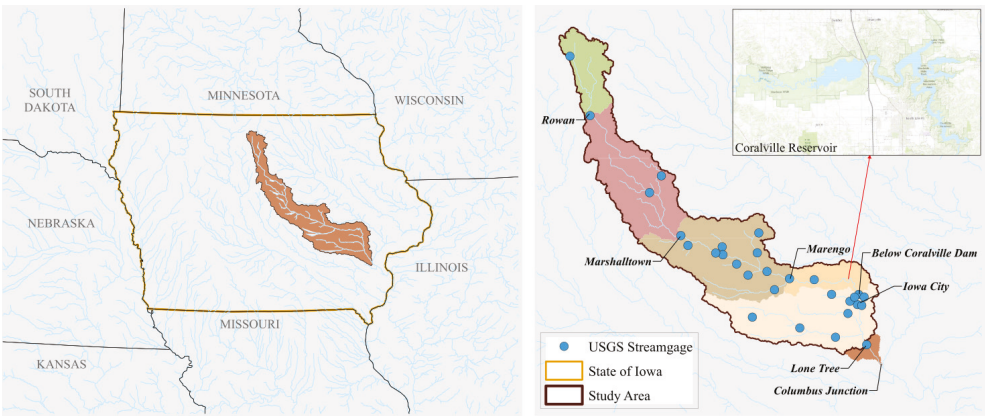


Figure 1. Iowa River Basin: locations of the USGS stream gauges and Coralville Reservoir.

A part of the Iowa River is regulated by the Coralville Dam within this basin, as indicated by its position in Figure 1. The dam has been operational since 1959, and it is an

earthfill dam spanning 426 m in length and standing 30 m tall. It creates the Coralville Reservoir on the Iowa River, situated 134 km upstream from its merging point with the Mississippi River and 8 km above Iowa City, IA. The main objective of this dam is to manage flood risks for regions along the Iowa River [34]. Above the dam, the Iowa River receives drainage from approximately 8068 km^2, primarily consisting of agricultural lands. Additionally, about 3052 km^2 of drainage areas from the dam location to Columbus Junction flow downstream, eventually reaching the Mississippi River.

2.2. Semi-Distributed HEC-HMS Hydrological Model

To simulate the rainfall–runoff transformation process in the basin, the semi-distributed event-based hydrologic modeling system (HEC-HMS) was employed [35]. Specifically, after delineating the Iowa River Basin using a 30-meter-resolution digital elevation model (DEM) and identifying the river network, rainfall data as a time series for each subbasin was fed into the model. The rainfall loss through infiltration was estimated using the Soil Conservation Service-Curve Number (SCS-CN) method [35]. This method calculates the direct runoff depth, denoted as P_e (mm)

$$P_e = \frac{(P - I_a)^2}{P - I_a + S} \tag{1}$$

where P is the accumulated rainfall depth (mm), and S represents the potential maximum retention (mm), linked to curve number (CN). Additionally, I_a denotes the initial abstraction (mm), which can be expressed as a function of S ($I_a = 0.2S$). The model inputs consist of I_a, CN, and the percentage of impervious area (no loss calculations are conducted on the impervious areas) [35].

The standard Clark Unit Hydrograph method [35] was employed to simulate runoff routing within the basin. The parameters essential for this transformation include the time of concentration, defined as the time needed for excess rainfall to travel from the most hydraulically remote point in the basin to the outlet, and the storage coefficient, which represents the attenuation resulting from storage effects throughout a basin. These parameters were calculated based on the physical properties of the basin. The recession method was also employed for determining the baseflow [35]. The parameters required for the recession method include the initial baseflow at the start of the simulation and the recession constant. The initial baseflow was specified using the discharge rate method, following the guidelines outlined in the HEC-HMS technical manual [35]. The recession constant was initially selected within the recommended range of 0.3–0.8 for each subbasin and was subsequently refined through model calibration to improve agreement with observed hydrographs.

Finally, the simulation of flood routing along natural channels was conducted using the Muskingum routing method [35]. This method uses the principle of conservation of mass to estimate the outflow hydrograph from the channel. Muskingum routing requires several parameters, including the Muskingum K parameter (representing travel time through a reach), the dimensionless coefficient X (ranging from 0.0 to 0.5, 0 for maximum attenuation and 0.5 for no attenuation), and the number of sub-reaches that can be calculated by dividing the Muskingum K parameter by the simulation time step.

The simulation of routing from a reservoir is also integrated within the HEC-HMS model through different methods: outflow structure routing, outflow curve routing, and specified release routing [35]. We utilized the outflow structure routing method. A dam reservoir was defined by the elevation–storage curve and by a specification of an

outlet. The outlet was defined by an inlet control with the discharge flowing out from the outlet when it is submerged, expressed as follows:

$$Q_c = A_c C_d \sqrt{2gh} \tag{2}$$

(where Q_c is the outlet discharge (cms), A_c represents the cross-sectional area of the outlet (m^2), which is circular with a diameter of 7 m in Coralville Dam [34], C_d denotes the discharge coefficient (equal to 0.8), g is the gravitational acceleration (9.81 m/s^2), and h stands for the total water head (m).

2.3. Model Calibration

The HEC-HMS model was calibrated using a single flood event that took place from 8 to 28 June 2008 within the study area. Hourly temporal resolution data from this event were utilized in the calibration process. This calibration involved providing streamflow observations obtained from USGS streamflow data at the outlet of each subbasin, corresponding to the locations of the stream gauges defined in Figure 1. For model calibration downstream of the Coralville Dam, streamflow data observed at a USGS station immediately downstream of the dam were used. This data included releases from the Coralville Dam, which were used as boundary conditions for the model in the downstream subbasin. In the calibration of the model, we considered statistical metrics such as the Nash–Sutcliffe efficiency (NSE), root mean squared error ratio (RSR), and percent bias (PBIAS), as formulated below.

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{N}(O_i - y_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \tag{3}$$

$$\text{RSR} = \frac{\text{RMSE}}{\sigma_o} = \frac{\sqrt{\sum_{i=1}^{N}(O_i - y_i)^2}}{\sqrt{\sum_{i=1}^{N}(O_i - \bar{O})^2}} \tag{4}$$

$$\text{PBIAS} = \frac{\sum_{i=1}^{N}(O_i - y_i) \times 100}{\sum_{i=1}^{N} O_i} \tag{5}$$

where O_i is the observed value, y_i is the simulated value, $\bar{O}$ is the average of observed values, σ_o is the standard deviation of observed values, and N is the total number of observations.

In general, model simulation can be judged as satisfactory if NSE > 0.50, RSR ≤ 0.70, and PBIAS is within $\pm 25\%$ for streamflow [36]. Table 1 presents the statistical metrics obtained after calibrating the HEC-HMS model for the study area. A peak streamflow is defined as a normalized streamflow value greater than 0.45, where the normalized streamflow is calculated by dividing the observed streamflow by the maximum streamflow recorded during the event. To further validate the calibrated model, we applied it to a distinct event that occurred from 26 April to 14 May 2013 within the studied basin. Table 2 presents the calculated metrics, all of which fall within these acceptable ranges. Therefore, the model was well-configured and ready for simulating flood hydrographs.

Table 1. Performance evaluation of the HEC-HMS model for the 8–28 June 2008 calibration event.

Gauge Station	NSE	RSR	PBIAS (%)	NSE (Peak)	RSR (Peak)	PBIAS (Peak) (%)
Rowan	0.90	0.30	−1.72	0.57	0.60	−1.40
Marshalltown	0.73	0.50	5.51	0.63	0.64	5.51
Marengo	0.87	0.40	4.23	0.75	0.53	3.50
Iowa City	0.96	0.20	−3.49	0.96	0.20	−3.22
Lone Tree	0.89	0.30	0.20	0.84	0.39	0.17

Table 2. Performance evaluation of the HEC-HMS model validated for the 26 April–14 May 2013 event.

Gauge Station	NSE	RSR	PBIAS (%)	NSE (Peak)	RSR (Peak)	PBIAS (Peak) (%)
Rowan	0.68	0.6	−8.72	0.54	0.72	−8.2
Marshalltown	0.64	0.6	−10.48	0.48	0.68	−9.82
Marengo	0.71	0.5	9.23	0.68	0.6	8.72
Iowa City	0.91	0.4	−5.36	0.90	0.4	−4.95
Lone Tree	0.62	0.6	6.30	0.47	0.78	5.64

3. Data Generation

3.1. Rainfall Event Generation

The stochastic storm transposition (SST) method [32] was utilized to create synthetic rainfall events that realistically represent the spatial and temporal variability in the study region. This approach extends the duration of extreme rainfall scenarios at a given location by transposing rainfall patterns from nearby areas exhibiting comparable extreme rainfall behavior.

The rainfall data for this study was generated using RainyDay [37], an open-source Python-based software available from the Hydroclimate Extremes Group (https://github.com/HydroclimateExtremesGroup/RainyDay, accessed on 20 September 2025). RainyDay integrates gridded rainfall data with the SST framework. The RainyDay implementation of the SST approach involves five primary stages: (1) define a transposition domain (A_D) that encompasses the target area (A_w). It is recommended that the A_D be at least four times larger than A_w [37]; (2) identify m temporally independent storms within A_D using n years of rainfall records. These storms are extracted based on accumulated rainfall over a duration t, with spatial characteristics (size, shape, and orientation) aligned to A_D to construct a "storm catalog"; (3) specify the number of storms per year, k, determined either via a Poisson distribution or an empirical distribution; (4) randomly sample k storms from the catalog; and (5) repeat steps 3 and 4 for a designated number of simulation years, $T_{\max}$, to produce $T_{\max}$ years of t-hour synthetic annual rainfall maxima for A_w.

For the present study, gridded Stage IV rainfall data spanning from 2002 to 2021, featuring a temporal resolution of 1 h and a spatial resolution of 4 km, were utilized to construct a storm catalog for spatiotemporal resampling. The A_D covered 95.36–90.11° W and 40.23–44.73° N. This extent is approximately four times larger than the study area, consistent with the recommendation. The storm catalog included the 200 highest-intensity rainfall events identified within the defined domain, which was considered sufficient to represent the basin's hydrometeorological variability while maintaining computational efficiency for subsequent simulations. These extreme events were selected based on 72-h rainfall accumulations evaluated over regions matching the Iowa River Basin in both size and shape. The random variable k was modeled using a Poisson distribution with a rate

parameter ($\lambda = 10$), calculated as the ratio of the total number of cataloged events to the number of years in the rainfall record.

In this study, the Iowa River Basin was divided into five subbasins before the Lone Tree, as illustrated in Figure 1. These subbasins were selected based on the presence of USGS gauge stations along the river's mainstream and the availability of data at those stations. Subsequently, the average rainfall time series for each subbasin was determined using zonal averaging.

3.2. Greedy Event Selection

While using the SST facilitates the generation of a wide range of rainfall events, enhancing both the diversity and duration of extreme precipitation, obtaining the corresponding streamflow responses necessitates hydrologic simulations. These simulations are computationally intensive and time-consuming, rendering the large-scale generation of rainfall–runoff events impractical. Moreover, random selection of rainfall events proves inefficient, often resulting in redundant or uninformative events to generate data that contribute minimally to model improvement. To address this limitation, active learning (AL) for regression offers a promising solution by reducing the number of required simulations through the strategic selection of the most informative rainfall events [38].

This paper focuses on pool-based AL [39], in which a pool of 200 rainfall events randomly generated using RainyDay is considered. The objective is to select a subset of these events for hydrological simulations to generate data, aiming to train an LSTM model that can provide highly accurate streamflow estimates for the remaining events. In contrast to the extensive literature on AL for classification problems, there are only a few approaches available for pool-based AL for regression [38–40]. The greedy sampling (GS) approach represents one of them. Several variants of GS have been proposed, including greedy sampling based on inputs (GSx), outputs (GSy), and a combination of both inputs and outputs (iGS), as outlined by Wu et al. [38].

In this study, we adopted the idea of GSx to enable a more informed and systematic selection of rainfall events. The GSx algorithm selects rainfall events from the pool of N events iteratively, assuming the pool with rainfall events represented by $\{E_n\}_{n=1}^{N}$, where E_n is a 72-dimensional vector corresponding to hourly rainfall values. Initially, GSx chooses the event closest to the centroid of all rainfall events as the first one, ensuring initial representativeness. Following this, each subsequent selection identifies the event with the maximum shortest Euclidean distance to previously chosen events, promoting diversity among the selected events. The goal is to guarantee that the selected events accurately represent the overall event distribution. Mathematically, the Euclidean distance between a candidate event E_n and each previously selected event E_m is calculated by Equation (6), where M is the total number of rainfall events that have been selected.

$$d_{nm} = \|E_n - E_m\| \qquad m = 1, \ldots, M \quad and \quad E_n \neq E_m \tag{6}$$

Then, the shortest distance is calculated for each candidate ($d_n = \min_m(d_{nm})$). Subsequently, the event with the maximum d_n is selected. Notably, the Euclidean distances computed using Equation (6) reflect differences among these rainfall patterns.

3.3. Data Generation Process

The goal of the greedy rainfall event selection using active learning was to ensure a diverse and representative subset of events within the overall event space. The selected rainfall events were subsequently fed into the HEC-HMS hydrological model, which was executed at an hourly temporal resolution. This process generated hourly streamflow data at computational points corresponding to USGS station locations in the Iowa River Basin.

The rainfall events were simulated sequentially, with each simulation spanning 20 days, during which rainfall occurred in the first 72 h. Additionally, we assumed that the dam's gates were fully open, with no initial water storage behind the dam.

In this study, we used datasets generated from the HEC-HMS model to train and evaluate LSTM models for streamflow prediction at the USGS station in Iowa City. The datasets consisted of two time-series features: rainfall and streamflow. The rainfall feature includes both historical and forecasted data. It consisted of hourly rainfall data from the subbasin covering Iowa City, with a total length of $t_h + t_f$, where t_h represents the number of past timesteps and t_f represents the maximum number of future timesteps. The streamflow feature consisted solely of historical data with a length of t_h. The ML models were designed to predict future streamflow with lead times up to t_f. For this study, we set $t_h = 60$ h and $t_f = 12$ h.

Each simulation of an individual rainfall event generated 480 data samples. By strategically selecting 20, 30, and 40 rainfall events through AL from a pool of 200, we generated three training datasets containing 9600, 14,400, and 19,200 samples, respectively. We refer to these training datasets as the 20-event, 30-event, and 40-event datasets. To create a test set, we randomly generated 20 additional rainfall events outside the original pool of 200, resulting in 9600 samples for model evaluation in this study.

4. Streamflow Prediction

This section presents the deterministic LSTM model developed for streamflow forecasting. We investigated the effect of training set size, generated through AL via the greedy selection of rainfall events.

4.1. LSTM-Based Encoder–Decoder

To manage sequential inputs and address issues such as exploding or vanishing gradients commonly encountered in traditional RNNs, LSTMs have been developed and extensively utilized across various fields [41]. In recent years, the domain of streamflow forecasting has also witnessed the application of LSTM models in different studies [14,15,20,42].

Figure 2 illustrates the basic architecture of an LSTM unit. Each LSTM unit typically has three gates: forget gate, input gate, and output gate. In the LSTM model, each time step comprises a specific component known as the cell state (C). The cell state preserves and conveys information pertinent to long-term memory within the model. The input sequence is presented as X, and the output is presented as h.

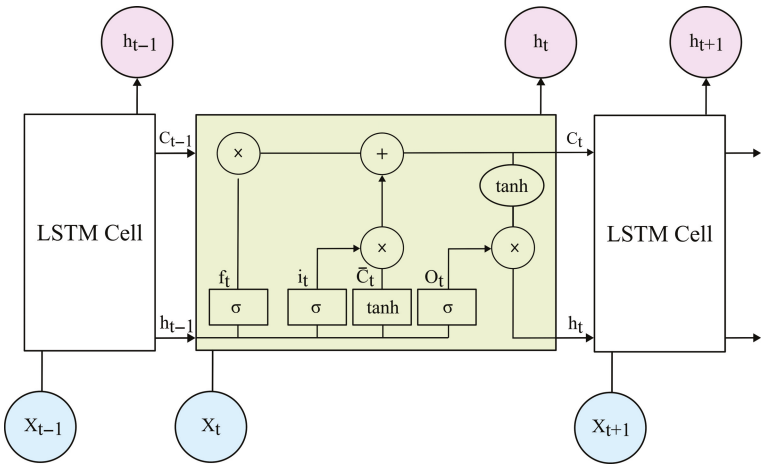


Figure 2. LSTM unit structure for the time step t to $t + 1$.

LSTM cells update six parameters per time step. Equations (7) to (12) present the details of the six parameters. The first parameter for each cell, which governs the degree to which the previous cell state should be forgotten, is the forget gate parameter f_t. In the given equations, weights are denoted by (W), and biases are denoted by (b). The input gate serves as the second parameter in an LSTM cell, playing a crucial role in deciding the new information to be added or retained in the cell state. $\bar{C}_t$ represents the candidate values for the new cell state. Subsequently, the cell state C_t is updated. Finally, the output parameter O_t is computed. The final gate is the output gate, which governs the extraction of information from the cell state and subsequently determines the current step h_t.

$$f_t = \sigma\,(W_f.[h_{t-1},\ X_t\,] + b_f) \tag{7}$$

$$i_t = \sigma\,(W_i.[h_{t-1},\ X_t\,] + b_i) \tag{8}$$

$$\bar{C}_t = tanh\,(W_C.[h_{t-1},\ X_t\,] + b_C) \tag{9}$$

$$C_t = f_t * C_{t-1} + i_t * \bar{C}_t \tag{10}$$

$$O_t = \sigma\,(W_O.[h_{t-1},\ X_t\,] + b_O) \tag{11}$$

$$h_t = O_t * tanh\,C_t \tag{12}$$

LSTM can address the issue of long-term dependencies, yet it is constrained by the requirement of having the same time steps for both input and output. Nonetheless, in contexts such as streamflow forecasting, as detailed in this paper, it is essential to include not just the rainfall data at the time steps we aim to predict but also the rainfall information from preceding hours. To address this issue, Sutskever et al. [43] introduced a neural network architecture known as encoder–decoder, enabling the model to operate with varying input and output time steps. As shown in Figure 3, the encoder LSTM for historical streamflow data in this study with t_h time steps produces a final output h_{t_h}, which is stored in a cell called the state vector. This state vector is then utilized as input for the decoder LSTM, which operates with t_f time steps.

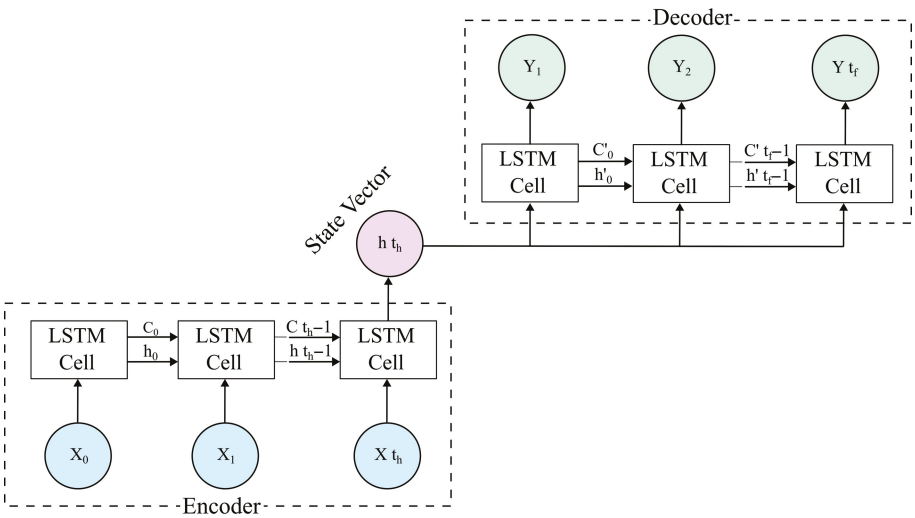


Figure 3. LSTM-based encoder–decoder model structure with t_h time step input and t_f time step output.

4.2. Model Design

In this study, we developed an LSTM-based encoder–decoder model, as illustrated in Figure 4, for hourly streamflow predictions up to 12 h into the future. The input features include historical and forecasted rainfall data as well as historical streamflow data. In this model framework, input time series are processed separately due to variations in their

lengths. The first LSTM encoder for the historical streamflow data has a length of $t_h = 60$, and the second LSTM encoder for the historical and forecasted rainfall data has a length of $t_h + t_f = 72$. In this section, we assume that future rainfall can be accurately forecasted. Later, in Section 6, we will account for rainfall forecast errors or uncertainties. Following the LSTM-based encoder–decoder architecture, several dense layers are employed to output forecasted streamflow for up to $t_f = 12$ h into the future.

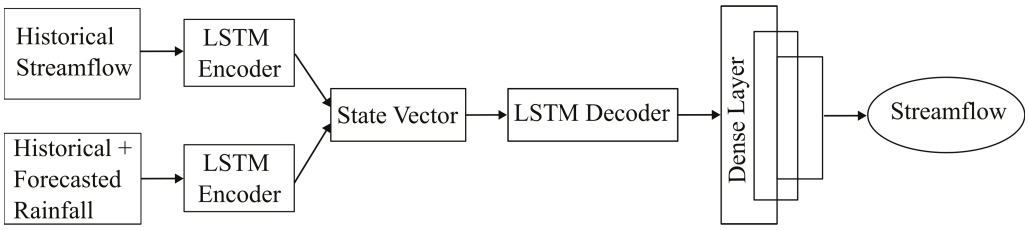


Figure 4. LSTM-based encoder–decoder model for streamflow forecasting.

We predominantly selected default settings and parameters for DL models available in TensorFlow-Keras. Each LSTM encoder layer contains 256 neurons, and the model includes a single LSTM decoder layer with 512 neurons. Following the decoder, there are six dense layers structured for time-series data, with neuron counts decreasing from 512 to 1. The ReLU activation function ($R(y) = max(0, y)$) is utilized in all dense layers. During the training phase, we implemented a dropout rate of 0.2 to mitigate overfitting and promote a sparse structure within the model. This dropout rate involved randomly deactivating 20 percent of the neural connections between nodes within the dense layers. The proposed DL model employs the mean squared error (MSE) as the loss function, defined as

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (Y_i - y_i)^2 \tag{13}$$

where N is the number of samples, and Y_i is the predicted value.

We employed the Adaptive Moment Estimation (Adam) optimization algorithm to minimize the loss function. The optimizer is initialized with a learning rate of 0.0001. During training, one out of every five instances in the dataset was allocated to the validation set after applying feature scaling using the MinMaxScaler. In this study, we performed model training, validation, and testing using Python (version 3.11.9), hosted on a machine with an Intel Core i7-12700K processor, an NVIDIA GeForce RTX 3070 Ti graphics card, and 64 GB of RAM.

4.3. Streamflow Prediction: Performance of Deterministic LSTMs

We evaluated the performance of deterministic LSTM models trained on three training datasets: the 20-event dataset (9600 samples), the 30-event dataset (14,400 samples), and the 40-event dataset (19,200 samples), as described in Section 3.3. An independent test set of 9600 samples was used to ensure a fair evaluation. Each LSTM model was trained following the procedure outlined in Section 4.2 and evaluated using the NSE metric as presented before (Equation (3)) with the difference that here we compare the HEC-HMS-simulated streamflow with the predicted streamflow generated by our developed model. The NSE in this context is defined as follows:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{N} (Y_i - y_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2} \tag{14}$$

where $\bar{y}_i$ denotes the average of simulated values.

Additionally, we trained a benchmark model on a larger dataset, referred to as the 200-event dataset, containing 96,000 samples generated from the original pool of 200 rainfall events. Figures 5 and 6 show the relationships between the LSTM model predictions and the HEC-HMS-simulated streamflow values on the test set at 1-h and 12-h lead times, respectively.

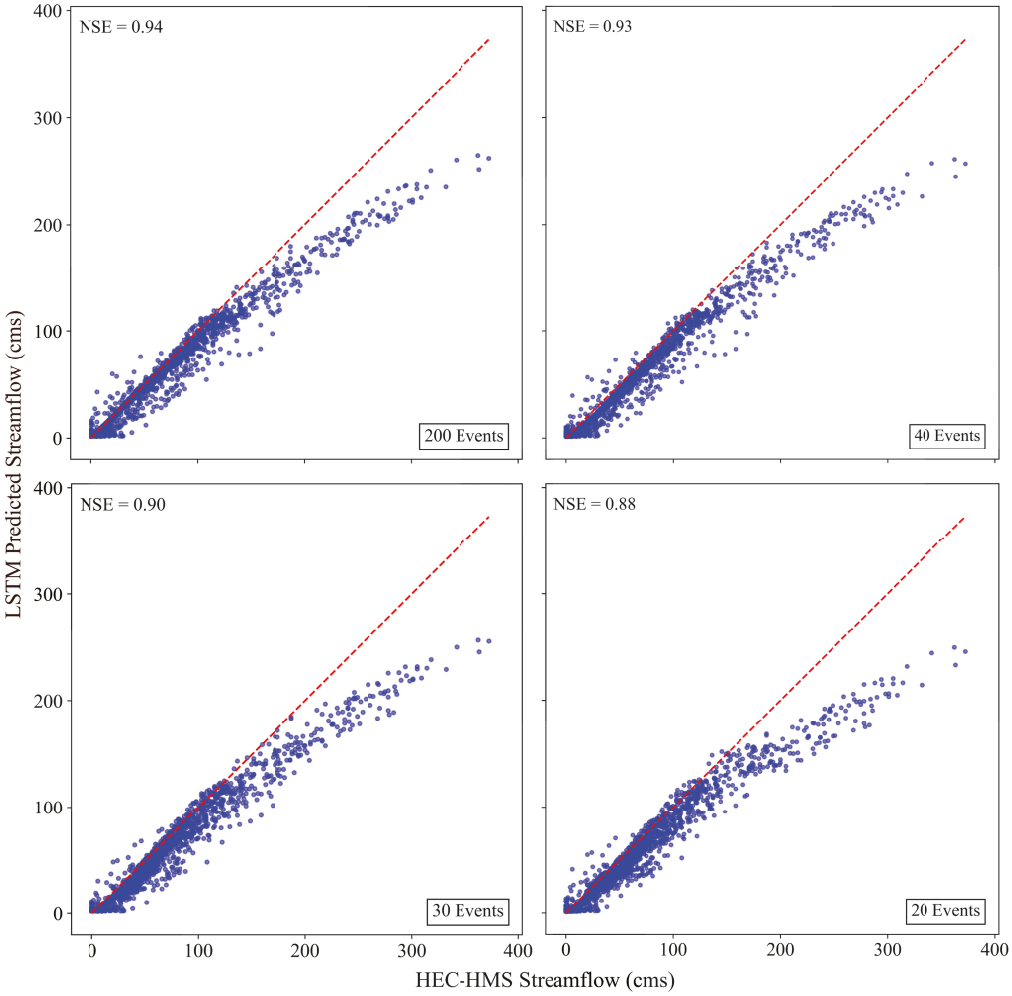


Figure 5. LSTM-predicted streamflow vs. HEC-HMS-simulated streamflow at a 1-h lead time for the test dataset across models trained with datasets of varying sizes. The red dashed line represents the 1:1 line, indicating perfect agreement.

Figure 5 illustrates that all models trained with different sizes of training sets have high predictive accuracy, especially for low streamflow values (less than 155 cms). The model assessment revealed very high NSE values of 0.94 and 0.93 between the predicted and HEC-HMS-simulated streamflow values from the LSTMs trained on the 200- and 40-event datasets, respectively. For two other models, trained on 30- and 20-event datasets, although plots show marginally lower NSE scores of 0.90 and 0.88 compared to the earlier ones, they still remain indicative of a strong correlation. For the predictions with a lead time of 12 h, Figure 6 depicts the same pattern as in Figure 5. However, the NSE values are slightly lower than those for the 1-h lead time predictions, with the model trained with the 200-event dataset outperforming the others.

We also calculated the NSEs to assess every LSTM across various lead times, as illustrated in Figure 7. Overall, all models achieved NSE values higher than 0.85, demonstrating strong predictive capabilities. The models performed exceptionally well at shorter lead times and maintained high accuracy up to 12 h ahead. In this context, the model trained

with the 200-event dataset outperformed the other three models. Although the model trained with the 40-event dataset did not reach the performance level of the 200-event model, it still achieved high NSE values across all lead times compared to models trained with fewer events, indicating its substantial predictive skill.

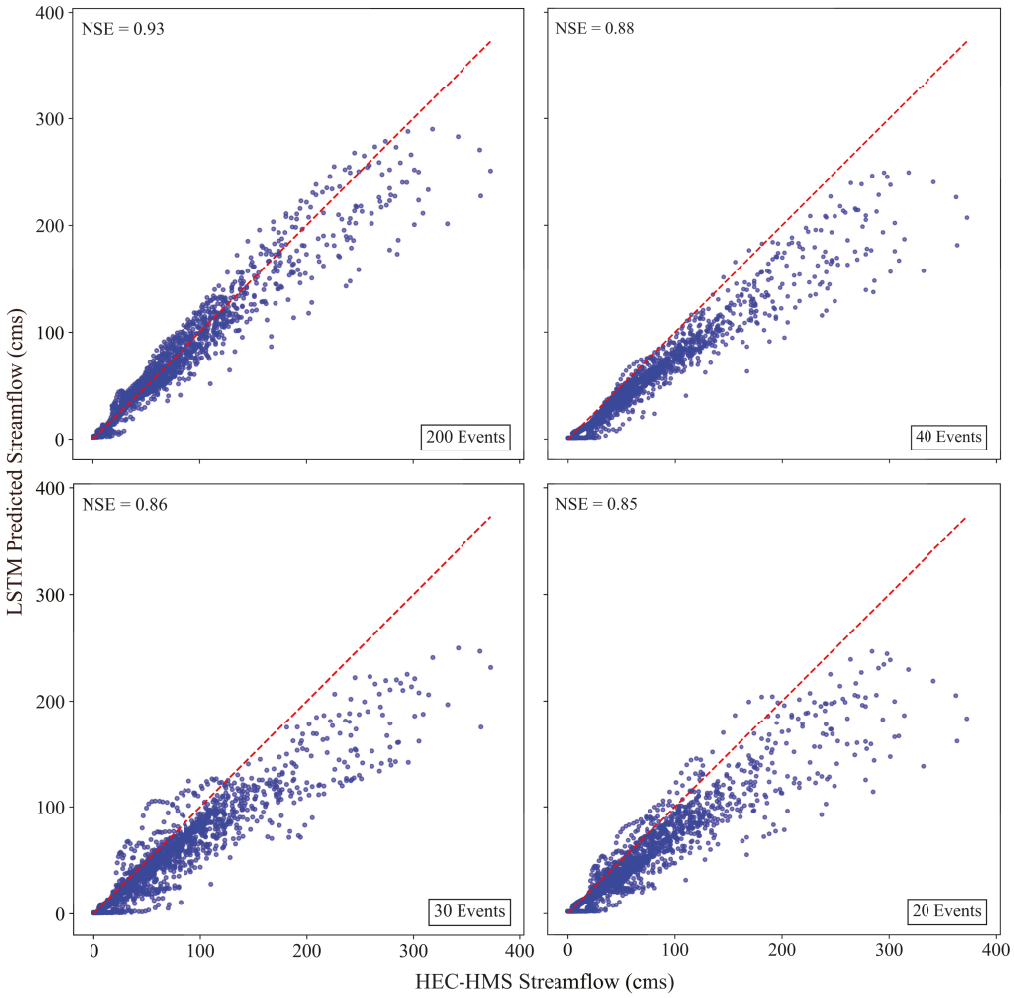


Figure 6. LSTM-predicted streamflow vs. HEC-HMS-simulated streamflow at a 12-h lead time for the test dataset across models trained with datasets of varying sizes.

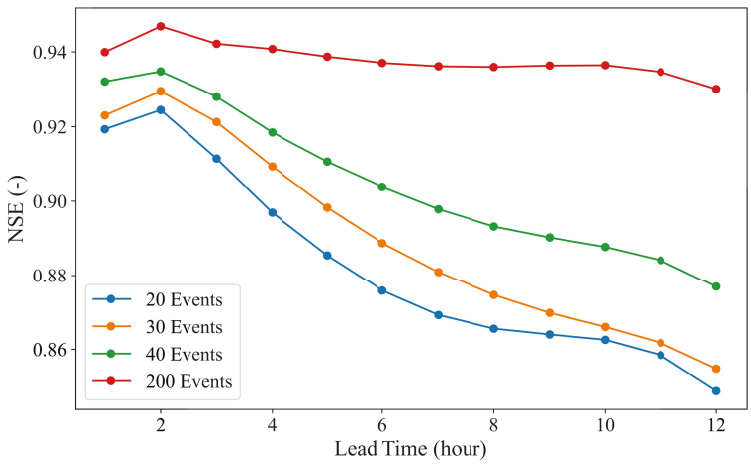


Figure 7. Performance of LSTM models trained with different training set sizes across all lead times.

These findings suggest that, while the 200-event dataset yields the highest accuracy, the 40-event dataset offers a practical alternative, striking an optimal balance between

performance and training efficiency. It provides sufficient data for effective training without overfitting and leads to better generalization and predictive performance. Consequently, the 40-event training set has been selected for further model improvement and analysis.

5. Streamflow Prediction with Emphasizing Peak Values

This section introduces two distinct loss functions designed to improve the performance of LSTMs in accurately capturing high streamflow values during peak flow events. Traditional loss functions, such as MSE, often struggle to emphasize extreme values, leading to the underestimation of peak flows. To address this limitation, we propose loss functions that prioritize high streamflow values, enhancing the model's ability to learn and predict extreme hydrological events more effectively.

5.1. Proposed LSTMs to Capture Higher Streamflow Values

In Figures 5 and 6, noticeable underestimations can be observed at higher streamflow values, a common issue in ML-based streamflow predictions when using MSE as a loss function [44]. Considering high streamflow values (>155 cms) only in the test set, the NSE could be as low as −0.90, meaning that the model prediction was worse than simply using the average value. This suggests that, while the overall accuracy was high, the ML model struggled with predicting higher streamflow values.

The MSE, as a symmetric loss function, treats all errors equally and emphasizes reducing the average error. This can disproportionately penalize larger errors, leading the model to prioritize minimizing errors for more frequent low streamflow values and potentially underestimate peak streamflow events. This underestimation is problematic in hydrological applications where accurately predicting peak values is crucial for flood warning systems and water resource management. To address this limitation, we introduce two asymmetric loss functions for our LSTM model to better capture higher streamflow values.

In contrast to symmetric loss functions, asymmetric loss functions are designed to handle cases where overestimations and underestimations have different consequences, thus assigning different penalties to these errors. This can improve the models' ability to predict higher values. This concept is rooted in several well-established practices, particularly in fields like economics, finance, and the power and energy sectors [45–48]. However, in the realm of hydrological forecasting, there are relatively fewer studies on this subject. We will highlight these studies in the following subsections.

5.1.1. LSTM with a Pinball Loss Function

When dealing with scenarios where the distribution of the target variable (the streamflow hydrograph) is non-Gaussian or skewed, the PL, also known as the quantile regression loss, can be used as an alternative [49]. This loss function assigns penalties to errors depending on the selected quantile. In the field of streamflow forecasting, the PL has been utilized in a limited number of studies within probabilistic frameworks for estimating the conditional quantiles of streamflow forecasts [30,50,51]. However, in our specific case, we have used the PL in a deterministic LSTM framework with the goal of enhancing the model's ability to predict peak streamflow events more accurately. We chose 0.8 and 0.9 quantiles, focusing on the upper end of the distribution to penalize underestimations. The PL function for a given quantile is defined as

$$\mathrm{PL}(y, Y) = \frac{1}{N}\sum_{i=1}^{N}(q \cdot (y_i - Y_i) \cdot \mathbf{1}(y_i > Y_i) + (1-q) \cdot (Y_i - y_i) \cdot \mathbf{1}(y_i \leq Y_i)) \tag{15}$$

where $\mathbf{1}(\cdot)$ is the indicator function, which is 1 if the condition inside the parentheses is true and 0 otherwise. q is the selected quantile. If choosing the 0.8 quantile ($q = 0.8$),

this function penalizes underestimations $Y_i < y_i$ much more heavily than overestimations $Y_i > y_i$, with weights of 0.8 and 0.2, respectively.

Figure 8 shows the performance of LSTM models using two different loss functions: PL and MSE. The comparisons are conducted for two different quantiles (0.8 and 0.9) and two different lead times (1 h and 12 h). Also, Table 3 presents the performance metrics for the entire test set, as well as for instances where the simulated streamflow values were higher than 155 cms. As we can see, in the case of $q = 0.9$, the model shows better overall performance (higher NSE) in capturing the variance of the simulated streamflow, with minimal scatter at the lead time of 1 h, particularly at higher streamflow values compared to MSE and PL with q = 0.8. This interpretation is consistent with the lead time of 12 h. However, as in Table 3, for streamflow values higher than 155 cms, the negative NSE values for a lead time of 12 h indicate that even the model with PL ($q = 0.9$) has difficulty capturing these values at longer lead times.

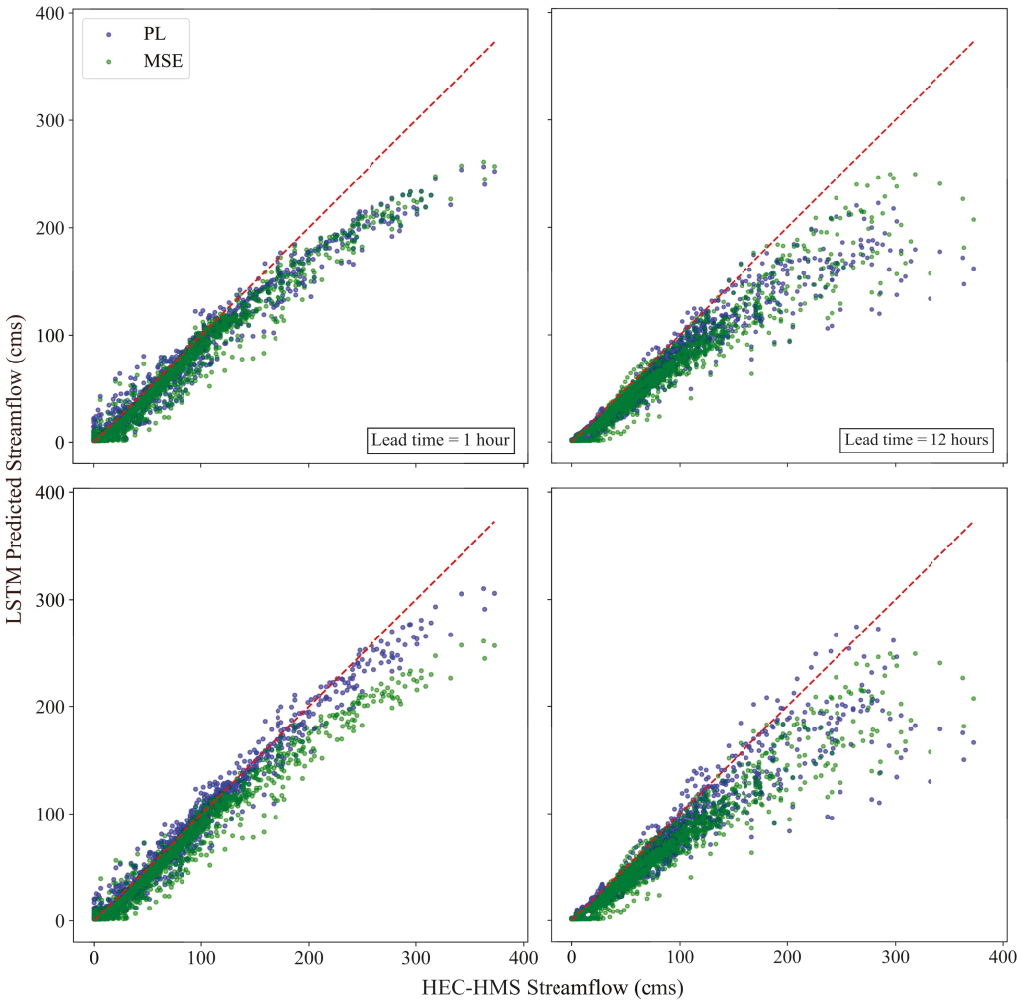


Figure 8. Performance of LSTMs using PL and MSE loss functions on the test set at 1-h (**left plots**) and 12-h lead times (**right plots**). The **upper plots** correspond to PL with $q = 0.8$, and the **lower plots** correspond to $q = 0.9$. The red dashed line represents the 1:1 line, indicating perfect agreement.

Table 3. Performance of LSTMs with PL and MSE loss functions at different lead times.

Loss Function	Lead Time (Hour)	NSE	NSE ($y_i > 155$ (cms))
PL (q = 0.8)	1	0.96	0.14
PL (q = 0.9)	1	0.98	0.82
MSE	1	0.93	0.12
PL (q = 0.8)	12	0.89	−1.33
PL (q = 0.9)	12	0.91	−0.91
MSE	12	0.88	−0.90

5.1.2. LSTM with an Asymmetric Peak Loss Function

We demonstrated that the PL function did not perform well in predicting high streamflow values at longer lead times. In this section, we developed an alternative approach, focusing more on improving the accuracy of peak value predictions. Our approach combined traditional loss functions (MSE) with additional terms that emphasize higher values. Specifically, we modified the MSE to be asymmetric by incorporating an additional penalty for errors when the simulated streamflow exceeds a certain threshold.

To implement this idea, we defined a peak threshold (T) and an asymmetry factor (F) in the new loss function and adjusted them to best suit our data and prediction goals. We named our implementation of this asymmetric loss function asymmetric peak (AP), which is defined as follows:

$$\mathrm{AP}(y, Y) = \frac{1}{N}\sum_{i=1}^{N}(y_i - Y_i)^2 + \frac{1}{N}\sum_{i=1}^{N}(y_i - Y_i)^2 \cdot F \cdot \mathbf{1}(y_i > T) \cdot \mathbf{1}(y_i > Y_i) \tag{16}$$

where $T = 0.45$ (normalized streamflow equal to 155 cms) and $F = 3.0$ were chosen based on model tuning.

Figure 9 shows the prediction performance of LSTM with the proposed loss function compared to the ones with the conventional MSE at the lead times of 1 h and 12 h. The figure clearly illustrates that the model with AP captured the higher streamflow values more effectively at both lead times, aligning with its purpose. Additionally, Table 4, representing the performance metrics of LSTM models with three different loss functions, highlights the strength of the LSTM with AP, even at the lead time of 12 h.

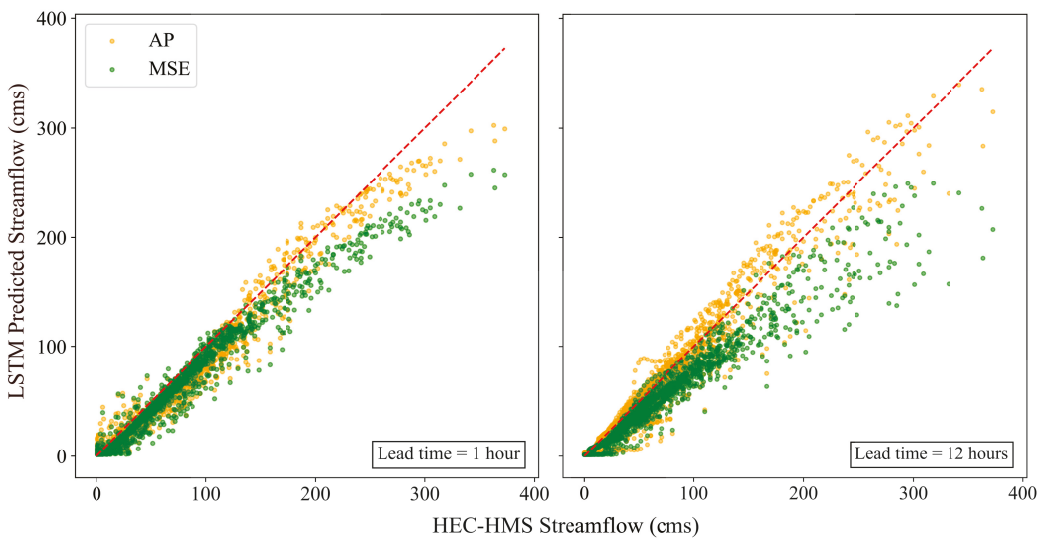


Figure 9. Performance of LSTM models using AP and MSE loss functions on the test set at two different lead times. The red dashed line represents the 1:1 line, indicating perfect agreement.

Table 4. Performance of LSTM models using AP and MSE loss functions for varying lead times.

Loss Function	Lead Time (Hour)	NSE	NSE ($y_i > T$)
AP	1	0.97	0.76
PL (q = 0.9)	1	0.98	0.82
MSE	1	0.93	0.12
AP	12	0.96	0.62
PL (q = 0.9)	12	0.91	−0.91
MSE	12	0.88	−0.90

To comprehensively compare the impacts of different loss functions across various lead times, Figure 10 is presented. It can be concluded that LSTMs with the proposed AP and the PL (q = 0.9) loss functions outperformed the LSTM model with the conventional MSE loss function at all lead times. Although the model with the PL loss function performed slightly better at shorter lead times (up to 3 h), the LSTM model with the proposed AP loss function demonstrated superior performance overall. In summary, the AP loss function emerged as the best performer for the LSTM model, consistently achieving high NSE values across all lead times and proving to be the most reliable choice for long-term predictions.

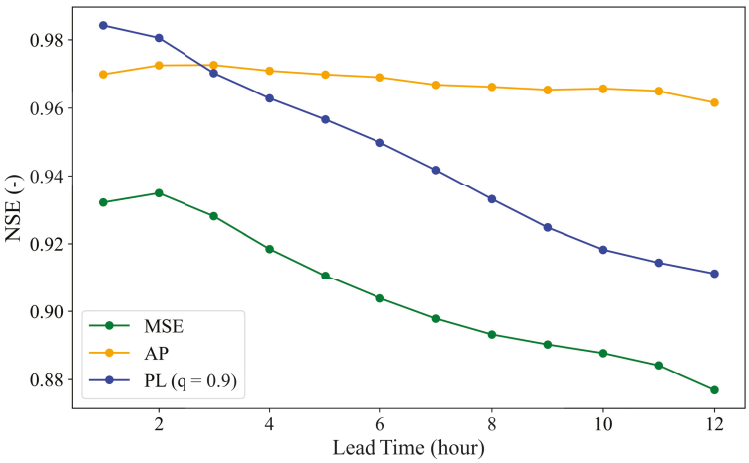


Figure 10. Performance of LSTM models using different loss functions across various lead times.

6. Streamflow Prediction with Rainfall Forecast Uncertainties

In the above studies, we have assumed that future rainfall forecasts are precise in our deterministic LSTM. However, this assumption does not hold in real-world applications, where rainfall forecast uncertainty is both substantial and often the primary source of uncertainty in streamflow forecasting. To address this issue, we developed a PLSTM for streamflow forecasting. In our approach, we first used archived rainfall forecast data along with subsequent actual observations to analyze the errors in rainfall forecasts. This uncertainty was then integrated into both the training and test datasets to better reflect real-world forecasting conditions. Utilizing our probabilistic LSTM model, we systematically quantified the uncertainty in streamflow predictions, extending up to 12 h into the future. By accounting for forecast uncertainty, our approach enhances the robustness and reliability of streamflow predictions, ultimately improving the model's ability to support early warning systems and flood risk management.

6.1. Rainfall Forecast Error Analysis

We analyzed the errors (uncertainties) in the forecasted rainfall over the next 12 h. To achieve this, we first obtained archived rainfall forecasts from January to November 2020 for the subbasin encompassing Iowa City from the High-Resolution Rapid Refresh (HRRR)

model. The HRRR, developed by NOAA, is a real-time convection-allowing atmospheric model with 3-kilometer spatial resolution and hourly updates, initialized on 3-kilometer grids incorporating radar data assimilation [52]. HRRR produces deterministic forecasts every hour out to 18 h for the continental United States (CONUS). We also obtained point observation data for the same period from the Iowa Environmental Mesonet [53]. Figure 11 illustrates the time series of 1 h into future rainfall forecasts versus the observed rainfall (top plot) at the Iowa City weather station. The corresponding time series of errors during the same period is also presented in Figure 11 (bottom plot).

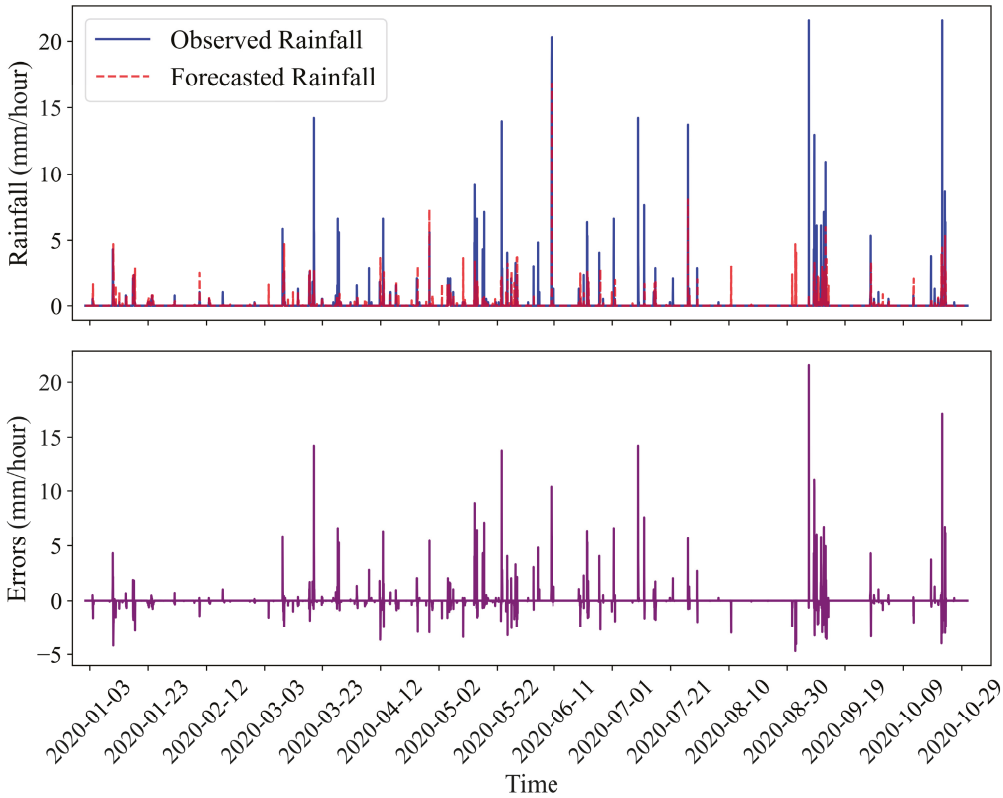


Figure 11. Time series of observed and forecasted rainfall (**top plot**) and forecast errors (**bottom plot**) from January to November 2020 at a 1-h lead time.

We defined errors as the difference between observed and forecasted rainfall, with negative errors indicating overestimations and positive errors implying underestimations. Although we have presented the results for just one-hour-ahead forecasts in this section, it is important to highlight that the average errors increased with lead times from 1 h to 12 h into the future.

We then stratified the observed rainfall values into five categories: no rainfall (rainfall $= 0$ (mm/hour), 6525 data points); light ($0 <$ rainfall < 1, 593 data points); moderate ($1 \leq$ rainfall < 5, 137 data points); heavy ($5 \leq$ rainfall < 10, 27 data points); and very heavy (rainfall ≥ 10, 14 data points). Subsequently, we fitted a distribution to the corresponding errors in each rainfall category. Figure 12 depicts the histogram of the errors and the fitted distributions for defined categories at the lead time of 1 h. It can be observed that underforecasting consistently occurs for heavy and very heavy rainfall.

Across all lead times up to 12 h, the best-fit probability distributions for the first three categories were normal distributions, whereas the "very heavy" category was best represented by a uniform distribution. For the fourth category, i.e., "heavy", the type of distribution varied with lead time, with both normal and uniform distributions being selected accordingly.

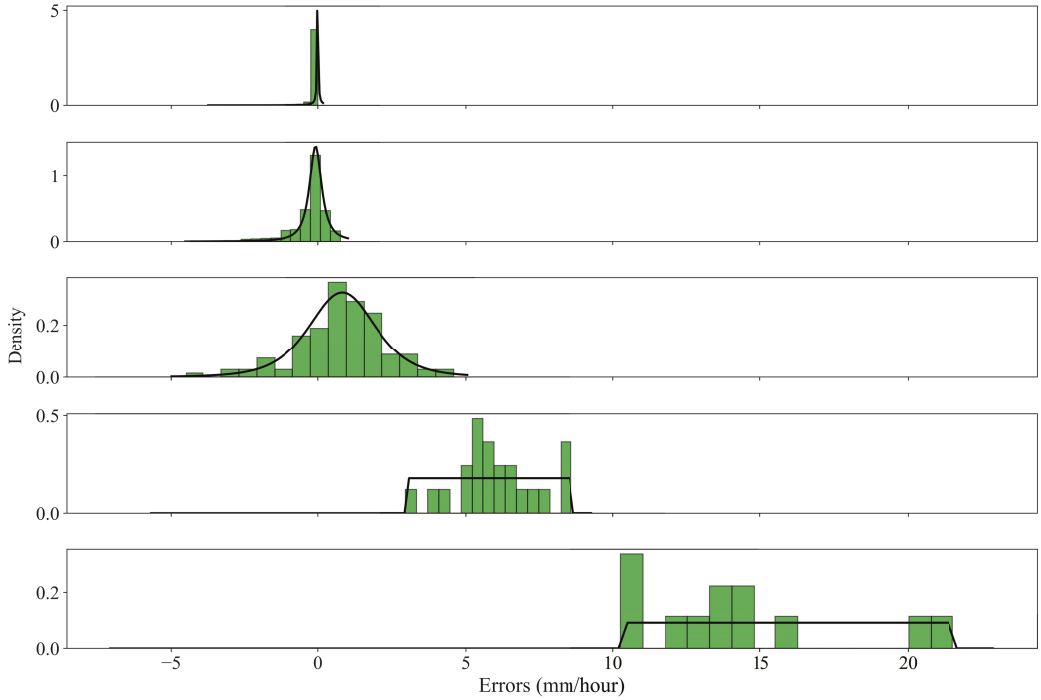


Figure 12. Histograms of rainfall forecast errors and their fitted probability distributions for different rainfall categories at 1-h lead time. The categories, arranged from **top** to **bottom**, are no rainfall, light, moderate, heavy, and very heavy.

6.2. Probabilistic LSTM

In previous sections, our discussion was centered on deterministic LSTM models, which provide point estimates of streamflow up to 12 h ahead. These models are limited in accounting for uncertainties, especially those in rainfall forecasts, and fail to offer a range of potential outcomes. Presenting a range of plausible scenarios is critical for informed and risk-aware decision-making. Therefore, advancing to PLSTM models is a necessary progression to address these constraints, enhancing the robustness and reliability of streamflow predictions.

We have constructed the PLSTM with NLL as the loss function, which involves training the model to directly optimize the log-likelihood of the observed data under the predicted probability distribution. This function's most general form is presented in Equation (17), which is minimized during the model training process.

$$\mathrm{NLL} = -\sum_{i=1}^{N} \log P(y_i|\theta) \tag{17}$$

Here, N is the number of samples, P is the probability density function (PDF) of the predicted distribution, and θ indicates the parameters of P.

Considering that streamflow values cannot be negative, we opted for the log-normal distribution. The probability density function of this distribution is defined as

$$P(y_i) = \frac{1}{y_i \sigma_i \sqrt{2\pi}} \exp\left(-\frac{(\ln y_i - \mu_i)^2}{2\sigma_i^2}\right) \tag{18}$$

where μ_i is the location parameter, and σ_i is the scale parameter.

Our PLSTM model uses the same input features as the deterministic LSTM, with separate LSTM encoders for historical streamflow and combined historical and forecasted rainfall inputs. However, in this study, we analyze the errors in rainfall forecasts, as discussed in Section 6.1, and introduce sampled noise into the rainfall data as input for the model. The encoder outputs are combined into a unified state vector, which is processed through an LSTM decoder and dense layers to predict the means and standard deviations of log-normal distributions, thereby providing confidence intervals for future streamflow predictions rather than single-point estimates.

This enables the model to capture both the expected values and associated uncertainties, providing confidence intervals rather than single-point estimates. The model closely follows the standard LSTM setup, but, as we mentioned above, it incorporates the NLL as the loss function. Additionally, the initial learning rate for the Adam optimizer is set to 0.0001.

6.3. Performance Evaluation Metrics

To evaluate the forecasts, we used established deterministic and probabilistic metrics from the hydrological modeling literature [24,51]. Reported metrics pertain to the test set, which was not used to train the model or tune hyperparameters.

Here, we employed NSE (Equation (14)), where the HEC-HMS-simulated streamflow is compared with the mean predicted streamflow generated by the PLSTM.

For evaluating probabilistic forecasts, we adopted three metrics: the continuous rank probability score (CRPS), the prediction interval coverage probability (PICP), and the mean prediction interval width (MPIW), as described below.

CRPS Equation (19) compares the forecasted and observed cumulative distribution functions (CDFs), providing a measure of how well the forecasted distribution matches the actual outcome. CRPS values range from 0 to infinity, with lower values indicating better prediction accuracy.

$$\text{CRPS} = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty}\left(F(Y_i) - \mathbf{1}(Y_i \geq y_i)\right)^2 dy \tag{19}$$

In this equation, F(Y_i) is the forecasted CDF.

PICP Equation (20) measures the proportion of observed data points that fall within the predicted intervals. A higher PICP indicates better reliability of the prediction intervals in capturing the true values. PICP values range from 0 to 1, where 1 indicates that all observed values fall within their respective prediction intervals.

$$\text{PICP} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(L_i \leq y_i \leq U_i) \tag{20}$$

L_i and U_i represent the lower and upper bounds of the prediction interval, respectively.

MPIW Equation (21) evaluates the average width of prediction intervals. It provides an indication of the uncertainty in predictions. Narrower intervals suggest higher confidence, while wider intervals indicate greater uncertainty.

$$\text{MPIW} = \frac{1}{N}\sum_{i=1}^{N}(U_i - L_i) \tag{21}$$

As discussed so far, the uncertainty in forecasted rainfalls up to 12 h into the future was integrated into the PLSTM model to output the predicted means and standard deviations of future streamflow values.

Figure 13 illustrates the comparison between HEC-HMS-simulated streamflow and predicted streamflow using the PLSTM model at 1-h and 12-h lead times. Both plots exhibit a general positive correlation, underscoring the model's proficiency in capturing streamflow trends. The shaded regions denote the 95 percent prediction interval, which expands at higher values of predicted streamflow, indicating increased uncertainty at these levels.

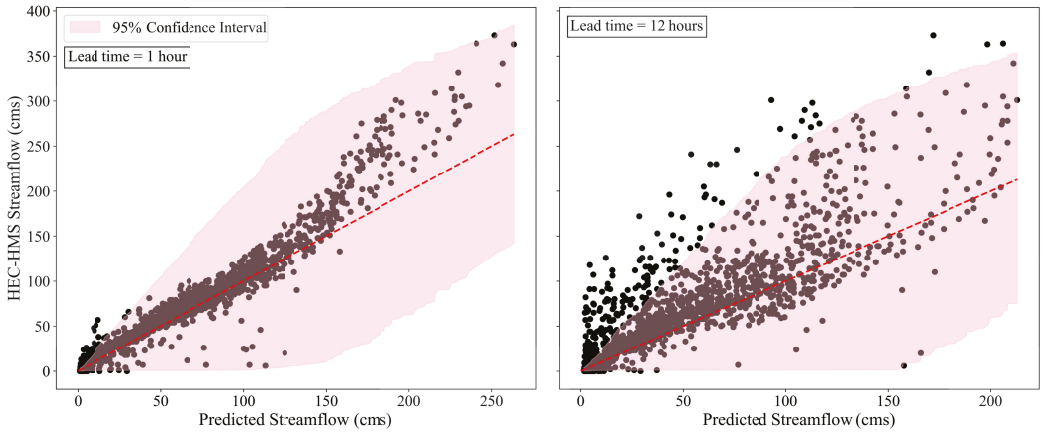


Figure 13. Performance of PLSTM at 1-h and 12-h lead times. Each black dot represents the mean predicted streamflow from the PLSTM model plotted against the corresponding streamflow simulated by HEC-HMS. The red dashed line represents the 1:1 line, indicating perfect agreement.

At the 1-h lead time, the model produces relatively narrow prediction intervals, particularly at lower streamflow values; the majority of the simulated streamflow values fall within these prediction intervals and closely align with the 1:1 line (red dashed line), indicating high prediction accuracy and confidence. Heavy rainfall usually leads to high streamflow. Consequently, prediction intervals widen at high streamflow values due to larger forecast errors associated with heavy and very heavy rainfall.

Conversely, the 12-h lead time plot reveals broader prediction intervals and a greater dispersion of data points around the 1:1 line, reflecting the increasing uncertainty associated with longer-term forecasts. Nonetheless, the prediction intervals continue to encompass a substantial portion of the simulated values, highlighting the model's capabilities even at extended lead times.

Also, Figure 14 presents point-wise predictions for several events in the test set at both 1-h and 12-h lead times, offering a clear visual assessment of the PLSTM model's predictive skill. The model effectively reproduces the shape and timing of streamflow peaks while maintaining well-calibrated probabilistic forecasts, with the majority of simulated streamflow values falling within the shaded 95 percent prediction intervals. As expected, the prediction intervals widen at the 12-h lead time, and discrepancies between the PLSTM mean predictions and the HEC-HMS-simulated streamflow—particularly around peak flows—become more pronounced. Overall, these results demonstrate the model's capability to reliably capture both the temporal dynamics and predictive uncertainty of streamflow across diverse hydrologic scenarios and lead times.

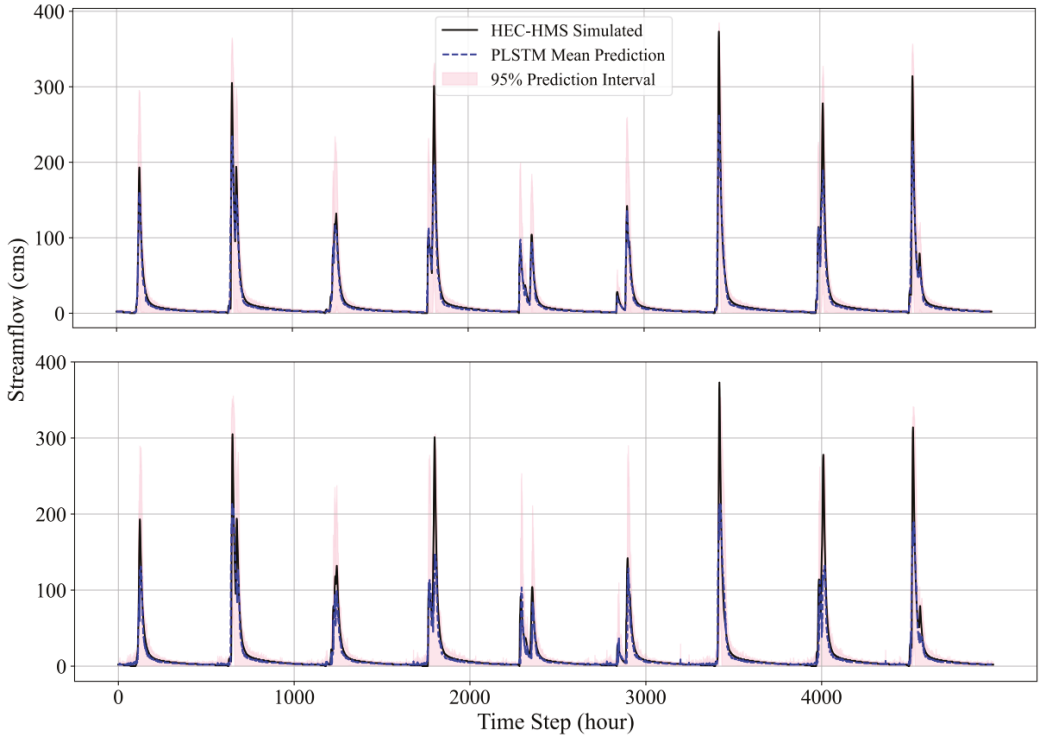


Figure 14. Comparison of streamflow time series from HEC-HMS simulations and PLSTM model predictions. The **upper plot** shows results at a 1-h lead time, while the **bottom plot** corresponds to a 12-h lead time.

Table 5 presents the performance metrics of the PLSTM model for lead times up to 12 h, complementing the insights from Figures 13 and 14. Note that higher values of NSE and PICP and lower values of MPIW and CRPS indicate better probabilistic model performance. The NSE for mean streamflow forecasts decreases from 0.9451 to 0.735 as the lead time increases from 1 to 12 h, highlighting a decline in predictive accuracy over longer lead times. The PICP values consistently remain above 0.9789, indicating that the 95 percent prediction intervals reliably encompass the true streamflow values across all the lead times. Although there is a slight decrease at longer lead times, the model continues to maintain good coverage. The MPIW values increase from 24.835 to 26.131 as the lead time extends from 1 to 12 h. This observed increase in MPIW signifies diminishing confidence in the model's predictions as lead time increases.

Table 5. Performance metrics of the PLSTM model at different lead times.

Lead Time (Hours)	NSE	PICP	MPIW	CRPS
1	0.9451	0.9956	24.8358	2.2604
2	0.9289	0.9958	25.574	2.3803
3	0.9182	0.9953	25.1313	2.5248
4	0.9003	0.9937	25.0292	2.7112
5	0.8790	0.9922	25.057	2.9270
6	0.8557	0.9906	25.1636	3.1485
7	0.8325	0.9884	25.3285	3.3621
8	0.8102	0.9864	25.5347	3.5644
9	0.7893	0.9847	25.7526	3.7521
10	0.7704	0.9832	25.9451	3.9245
11	0.7532	0.9809	26.0816	4.0781
12	0.7352	0.9789	26.1311	4.2355

7. Conclusions and Outlook

This study proposes three advanced ML techniques to enhance data-driven streamflow forecasting. First, AL was employed to optimally select rainfall events, minimizing the efforts of hydrological modeling and simulations required for data generation. The LSTM trained with 40 selected rainfall events from a pool of 200 randomly generated events achieved NSE values of 0.93 and 0.88 at 1-h and 12-h lead times, respectively—only marginally lower (by less than 1% and 5%) than the model trained with all 200 events. This finding demonstrates that a substantially smaller yet strategically selected training set can preserve predictive accuracy comparable to that of a fully trained model. Overall, this approach effectively identified the most informative rainfall events for model training while markedly reducing computational cost and data generation time.

Secondly, we introduced a novel loss function tailored specifically for peak streamflow prediction. By incorporating the developed asymmetric peak (AP) loss function, the model achieved substantial quantitative improvements over the conventional MSE loss. At a 1-h lead time, the AP-based model improved the overall NSE from 0.93 to 0.97 (a 4.3% increase) and peak-flow NSE from 0.12 to 0.76—representing more than a sixfold improvement in capturing high-flow values. Similarly, at a 12-h lead time, the AP loss enhanced the overall NSE from 0.88 to 0.96 (a 9.1% increase) and peak-flow NSE from –0.90 to 0.62, demonstrating a complete reversal from poor to highly reliable peak predictions. These findings confirm that AP loss delivers superior skill in capturing high streamflow values, effectively addressing one of the most critical challenges in hydrological forecasting. Additionally, the developed loss function outperformed the state-of-the-art PL function and holds promise for applications in ML models in other domains, such as economics and finance.

Lastly, we developed and implemented a probabilistic LSTM model that effectively incorporates rainfall forecast errors into streamflow predictions. The model achieved PICP higher than 0.97, NSE decreasing from 0.94 to 0.73, CRPS increasing from 2.26 to 4.24, and MPIW increasing from 24.835 to 26.131 across lead times up to 12 h, offering well-calibrated uncertainty quantification while maintaining high predictive skill. These contributions collectively improve the accuracy, reliability, and practical utility of streamflow forecasts, providing valuable insights for decision-making and flood management efforts.

While this research primarily focused on streamflow forecasting, the proposed methods can be applied to predict crucial variables in various related fields, including meteorology (wind speed and precipitation), environmental science (nitrate concentration), water resource management (reservoir levels and water demand), and agriculture (evapotranspiration).

In the current study, only one outflow configuration (gates fully open and initially zero storage) was considered. Building on the strengths of our approach, future research will explore additional configurations. While rainfall events were limited to 3 days for data generation in this study, the approach can be easily extended to events of varying lengths. Additionally, to improve the performance of our PLSTM model at longer lead times, we are considering investigating the effect of combining NLL and AP loss functions. These directions aim to refine our models further and enhance their applicability across diverse scenarios. Furthermore, alternative architectures, such as GRU, CNN, and Transformer networks, will be explored for comparative analysis and potential model improvement.

Author Contributions: Conceptualization, S.T., R.M., F.G. and S.X.; methodology, S.T. and S.X.; software, S.T.; validation, S.T.; formal analysis, S.T.; investigation, S.T. and S.X.; resources, S.T., R.M., F.G. and S.X.; data curation, S.T.; writing—original draft preparation, S.T.; writing—review and editing, S.T., R.M., F.G. and S.X.; visualization, S.T.; supervision, R.M. and S.X.; project administration, S.X.; funding acquisition, S.X. All authors have read and agreed to the published version of the manuscript.

Funding: This material is based upon work supported by the National Science Foundation under Grant Number 2226936 and the U.S. Department of Education under Grant Number ED#P116S210005. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation and the U.S. Department of Education.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting this study are available from the author's GitHub repository at https://github.com/soheylam/LSTM_PLSTM_ActiveLearning (accessed on 28 October 2025).

Acknowledgments: During the preparation of this manuscript/study, the author(s) used GPT-3.5 and GPT-4 for the purposes of improving the readability and language. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ML	Machine learning
SST	Stochastic storm transposition
LSTM	Long short-term memory
AI	Artificial intelligence
SVMs	Support vector machines
DL	Deep learning
RNNs	Recurrent neural networks
SWAT	Soil water assessment tool
AL	Active learning
PL	Pinball loss
PLSTM	Probabilistic long short-term memory
NLL	Negative log-likelihood
HMS	Hydrologic modeling system
DEM	Digital elevation model
SCS	Soil conservation service
CN	Curve number
NSE	Nash–Sutcliffe efficiency
RSR	Root mean squared error ratio
PBIAS	Percent bias
GS	Greedy sampling
AP	Asymmetric peak
CRPS	Continuous rank probability score
PICP	Prediction interval coverage probability
MPIW	Mean prediction interval width

References

1. Zhao, X.; Lv, H.; Wei, Y.; Lv, S.; Zhu, X. Streamflow forecasting via two types of predictive structure-based gated recurrent unit models. *Water* **2021**, *13*, 91. [CrossRef]
2. Nguyen, H.D.; Pham Van, C.; Nguyen, Q.-H.; Bui, Q.-T. Daily streamflow prediction based on the long short-term memory algorithm: A case study in the Vietnamese Mekong Delta. *J. Water Clim. Change* **2023**, *14*, 1247–1267. [CrossRef]
3. Farfán-Durán, J.F.; Cea, L. Streamflow forecasting with deep learning models: A side-by-side comparison in Northwest Spain. *Earth Sci. Inform.* **2024**, *17*, 5289–5315. [CrossRef]

4. Yifru, B.A.; Lim, K.J.; Lee, S. Enhancing streamflow prediction physically consistently using process-based modeling and domain knowledge: A review. *Sustainability* **2024**, *16*, 1376. [CrossRef]
5. Zhang, J.; Chen, X.; Khan, A.; Zhang, Y.; Kuang, X.; Liang, X.; Taccari, M.; Nuttall, J. Daily runoff forecasting by deep recursive neural network. *J. Hydrol.* **2021**, *596*, 126067. [CrossRef]
6. Sarker, S.; Leta, O.T. Review of watershed hydrology and mathematical models. *Eng* **2025**, *6*, 129. [CrossRef]
7. Paniconi, C.; Lauvernet, C.; Rivard, C. Exploration of coupled surface–subsurface hydrological model responses and challenges through catchment- and hillslope-scale examples. *Front. Water* **2025**, *7*, 1553578. [CrossRef]
8. Xia, Q.; Fan, Y.; Zhang, H.; Jiang, C.; Wang, Y.; Hua, X.; Liu, D. A review on the development of two-way coupled atmospheric-hydrological models. *Sustainability* **2023**, *15*, 2803. [CrossRef]
9. Jahanbani, H.; Ahmed, K.; Gu, B. Data-driven artificial intelligence-based streamflow forecasting: A review of methods, applications, and tools. *JAWRA J. Am. Water Resour. Assoc.* **2024**, *60*, 1095–1119. [CrossRef]
10. Granata, F.; Gargano, R.; De Marinis, G. Support vector regression for rainfall-runoff modeling in urban drainage: A comparison with the EPA's storm water management model. *Water* **2016**, *8*, 69. [CrossRef]
11. Yan, J.; Jin, J.; Chen, F.; Yu, G.; Yin, H.; Wang, W. Urban flash flood forecast using support vector machine and numerical simulation. *J. Hydroinform.* **2018**, *20*, 221–231. [CrossRef]
12. Shi, T.; Shide, K. A comparative analysis of LSTM, GRU, and Transformer models for construction cost prediction with multidimensional feature integration. *J. Asian Archit. Build. Eng.* **2025**, 1–16. [CrossRef]
13. Xiao, J.; Deng, T.; Bi, S. Comparative analysis of LSTM, GRU, and Transformer models for stock price prediction. In Proceedings of the International Conference on Digital Economy, Blockchain and Artificial Intelligence, Guangzhou, China, 23–25 August 2024; pp. 103–108.
14. Kratzert, F.; Klotz, D.; Brenner, C.; Schulz, K.; Herrnegger, M. Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 6005–6022. [CrossRef]
15. Xiang, Z.; Yan, J.; Demir, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour. Res.* **2020**, *56*, e2019WR025326. [CrossRef]
16. Guo, Y.; Yu, X.; Xu, Y.; Chen, H.; Gu, H.; Xie, J. AI-based techniques for multi-step streamflow forecasts: Application for multi-objective reservoir operation optimization and performance assessment. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 5951–5979. [CrossRef]
17. Yin, H.; Wang, F.; Zhang, X.; Zhang, Y.; Chen, J.; Xia, R.; Jin, J. Rainfall-runoff modeling using long short-term memory based step-sequence framework. *J. Hydrol.* **2022**, *610*, 127901. [CrossRef]
18. Luppichini, M.; Barsanti, M.; Giannecchini, R.; Bini, M. Deep learning models to predict flood events in fast-flowing watersheds. *Sci. Total Environ.* **2022**, *813*, 151885. [CrossRef]
19. Hunt, K.M.R.; Matthews, G.R.; Pappenberger, F.; Prudhomme, C. Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 5449–5472. [CrossRef]
20. Arsenault, R.; Martel, J.; Brunet, F.; Brissette, F.; Mai, J. Continuous streamflow prediction in ungauged basins: Long short-term memory neural networks clearly outperform traditional hydrological models. *Hydrol. Earth Syst. Sci.* **2023**, *27*, 139–157. [CrossRef]
21. Chen, S.; Huang, J.; Huang, J. Improving daily streamflow simulations for data-scarce watersheds using the coupled SWAT-LSTM approach. *J. Hydrol.* **2023**, *622*, 129734. [CrossRef]
22. Tursun, A.; Xie, X.; Wang, Y.; Liu, Y.; Peng, D.; Rusuli, Y.; Zheng, B. Reconstruction of missing streamflow series in human-regulated catchments using a data integration LSTM model. *J. Hydrol. Reg. Stud.* **2024**, *52*, 101744. [CrossRef]
23. Klotz, D.; Kratzert, F.; Gauch, M.; Sampson, A.K.; Brandstetter, J.; Klambauer, G.; Hochreiter, S.; Nearing, G. Uncertainty estimation with deep learning for rainfall–runoff modeling. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 1673–1693. [CrossRef]
24. Jahangir, M.S.; Quilty, J. Generative deep learning for probabilistic streamflow forecasting: Conditional variational auto-encoder. *J. Hydrol.* **2024**, *629*, 130498. [CrossRef]
25. Tran, V.N.; Dwelle, M.C.; Sargsyan, K.; Ivanov, V.Y.; Kim, J. A novel modeling framework for computationally efficient and accurate real-time ensemble flood forecasting with uncertainty quantification. *Water Resour. Res.* **2020**, *56*, e2019WR025727. [CrossRef]
26. Delottier, H.; Doherty, J.; Brunner, P. Data space inversion for efficient uncertainty quantification using an integrated surface and sub-surface hydrologic model. *Geosci. Model Dev.* **2023**, *16*, 4213–4231. [CrossRef]
27. Troin, M.; Arsenault, R.; Wood, A.W.; Brissette, F.; Martel, J. Generating ensemble streamflow forecasts: A review of methods and approaches over the past 40 years. *Water Resour. Res.* **2021**, *57*, e2020WR028392. [CrossRef]
28. Hauswirth, S.M.; Bierkens, M.F.P.; Beijk, V.; Wanders, N. The suitability of a seasonal ensemble hybrid framework including data-driven approaches for hydrological forecasting. *Hydrol. Earth Syst. Sci.* **2023**, *27*, 501–517. [CrossRef]
29. Hao, Y.; Baik, J.; Tran, H.; Choi, M. Quantification of the effect of hydrological drivers on actual evapotranspiration using the Bayesian model averaging approach for various landscapes over Northeast Asia. *J. Hydrol.* **2022**, *607*, 127543. [CrossRef]

30. Jahangir, M.S.; You, J.; Quilty, J. A quantile-based encoder-decoder framework for multi-step ahead runoff forecasting. *J. Hydrol.* **2022**, *619*, 129269. [CrossRef]
31. Haddad, K. A comprehensive review and application of Bayesian methods in hydrological modelling: Past, present, and future directions. *Water* **2025**, *17*, 1095. [CrossRef]
32. Gurbuz, F.; Mudireddy, A.; Mantilla, R.; Xiao, S. Using a physics-based hydrological model and storm transposition to investigate machine-learning algorithms for streamflow prediction. *J. Hydrol.* **2024**, *628*, 130504. [CrossRef]
33. Alabbad, Y.; Demir, I. Comprehensive flood vulnerability analysis in urban communities: Iowa case study. *Int. J. Disaster Risk Reduct.* **2022**, *74*, 102955. [CrossRef]
34. U.S. Army Corps of Engineers, Rock Island District. *Coralville Lake Water Control Plan Update Report with Integrated Environmental Assessment*; U.S. Army Corps of Engineers: Rock Island, IL, USA, 2022.
35. U.S. Army Corps of Engineers, Hydrologic Engineering Center. HEC-HMS User's Manual, Version 4.13. Hydrol Eng Center: Davis, CA, USA. Available online: https://www.hec.usace.army.mil/confluence/hmsdocs/hmsum/latest (accessed on 11 October 2025).
36. Moriasi, D.; Arnold, J.G.; Liew, M.W.V.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **2007**, *50*, 885–900. [CrossRef]
37. Wright, D.B. RainyDay: Rainfall Hazard Analysis System. Available online: https://github.com/HydroclimateExtremesGroup/RainyDay (accessed on 20 September 2025).
38. Wu, D.; Lin, C.; Huang, J. Active learning for regression using greedy sampling. *Inf. Sci.* **2019**, *474*, 90–105. [CrossRef]
39. Bi, J.; Xu, Y.; Conrad, F.; Wiemer, H.; Ihlenfeldt, S. A comprehensive benchmark of active learning strategies with AutoML for small-sample regression in materials science. *Sci. Rep.* **2025**, *15*, 37167. [CrossRef]
40. Chen, Y.; Deierling, P.; Xiao, S. Exploring active learning strategies for predictive models in mechanics of materials. *Appl. Phys. A* **2024**, *130*, 588. [CrossRef]
41. Waqas, M.; Humphries, U.W. A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX* **2024**, *13*, 102946. [CrossRef] [PubMed]
42. Ni, L.; Wang, D.; Singh, V.P.; Wu, J.; Wang, Y.; Tao, Y.; Zhang, J. Streamflow and rainfall forecasting by two long short-term memory-based models. *J. Hydrol.* **2020**, *583*, 124296. [CrossRef]
43. Sutskever, I.; Vinyals, O.; Le Q.V. Sequence to sequence learning with neural networks. In Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 3104–3112.
44. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [CrossRef]
45. Dessain, J. Improving the Prediction of Asset Returns with Machine Learning by Using a Custom Loss Function. *Adv. Artif. Intell. Mach. Learn.* **2023**, *3*, 1640–1653. [CrossRef]
46. Wang, Y.; Gan, D.; Sun, M.; Zhang, N.; Lu, Z.; Kang, C. Probabilistic individual load forecasting using pinball loss guided LSTM. *Appl. Energy* **2019**, *235*, 10–20. [CrossRef]
47. Wu, J.; Wang, Y.; Tian, Y.; Burrage, K.; Cao, T. Support vector regression with asymmetric loss for optimal electric load forecasting. *Energy* **2021**, *223*, 119969. [CrossRef]
48. Zhang, J.; Wang, Y.; Hug, G. Cost-oriented load forecasting. *Electr. Power Syst. Res.* **2022**, *205*, 117723. [CrossRef]
49. Tagasovska, N.; Lopez-Paz, D. Single-model uncertainties for deep learning. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 6417–6428.
50. Tyralis, H.; Papacharalampous, G. Quantile-based hydrological modelling. *Water* **2022**, *13*, 3420. [CrossRef]
51. Cao, C.; He, Y.; Cai, S. Probabilistic runoff forecasting considering stepwise decomposition framework and external factor integration structure. *Expert Syst. Appl.* **2024**, *236*, 121350. [CrossRef]
52. NOAA. HRRR: High-Resolution Rapid Refresh. Available online: https://rapidrefresh.noaa.gov/hrrr/ (accessed on 11 October 2025).
53. Iowa Environmental Mesonet. ASOS/AWOS Hourly Precipitation Data. Available online: https://mesonet.agron.iastate.edu/request/asos/hourlyprecip.phtml (accessed on 11 October 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

How Does Short Video Advertisement Congruence Drive Sales? The Underlying Mechanism of Sociability

Dongmei Han [1,2,3], Wangyan Jin [1,*], Zhengze Wu [1] and Ruyi Ge [2]

[1] School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200082, China; dongmeihan@mail.shufe.edu.cn (D.H.); zhengze_wu@foxmail.com (Z.W.)
[2] Faculty of Business Information, Shanghai Business School, Shanghai 200235, China; gery@sbs.edu.cn
[3] Shanghai Key Laboratory of Financial Information Technology, Shanghai University of Finance and Economics, Shanghai 200433, China
* Correspondence: wangyanjin@stu.sufe.edu.cn

Abstract

With the rapid growth of social media platforms, short video advertisements (SVAs) have been a dominant channel for product sales. However, how to design SVAs that effectively drive product sales, especially in relation to previous SVAs and the related product information, remains underexplored. This study investigates how SVA title congruence influences sales performance through the mediating role of sociability. Specifically, we conceptualized video-video title congruence and video-product title congruence as two forms of content congruence and investigated their effects using data collected from Douyin, the leading short video platform. The empirical results with two-way fixed effects show that high video-video title congruence and low video-product title congruence are both associated with higher product sales. Sociability mediates the relationship between title congruence and sales performance. This study also finds that the creation frequency and product brand significantly moderate these relationships. Furthermore, this study develops several checks to ensure the robustness of the research model and findings, including the Heckman two-stage test. These findings provide theoretical insights into content creation strategies and offer practical implications for both creators and platform managers.

Keywords: congruence; short video advertisements (SVAs); sociability; product sales

1. Introduction

Short video advertisements (SVAs) have become a dominant force in social commerce, transforming how consumers discover and purchase products. These SVAs, lasting from a few seconds to several minutes, leverage the interactive and emotionally engaging nature of social media to promote products [1]. They often contain embedded purchase links that enable consumers to complete instant transactions while watching the video [2]. Due to their informativeness, rapid dissemination, and low production cost [3], SVAs significantly enhance consumer engagement [4] and influence online purchasing decisions [5,6]. The proliferation of short videos has reshaped online consumption behavior, yet it has also raised concerns such as content saturation, user fatigue, and declining trust in created content. This situation highlights a pressing social issue: How to balance commercial persuasion with authentic social interaction in digital environments.

Unlike traditional e-commerce platforms (i.e., Taobao and JD, Douyin (the Chinese version of TikTok) has become China's leading short video platform, integrating entertainment, social interaction, and e-commerce into a unified digital ecosystem [7,8]. Compared

with competing platforms such as Kwai in China and its international counterpart TikTok, Douyin distinguishes itself through its sophisticated recommendation algorithms, highly personalized content feeds, and seamless in-app purchasing experience that effectively transforms user engagement into sales [9]. This integration enables both businesses and individuals to promote and sell products through short video advertisements (SVAs) [5,10]. However, it also faces challenges such as content homogenization, algorithmic dependency, and intensified competition for visibility [11,12]. Within this environment, creators can obtain valuable insights into consumer preferences and engagement intensity through sociability metrics, such as likes, comments, and shares [13]. Although some viewers may purchase products directly without engaging in social interactions, these social performance indicators not only reflect a proportion of audience acceptance and content popularity but also influence the platforms' recommendation algorithms, which, in turn, shape visibility, conversion, and sales performance. Therefore, determining which SVA features can enhance sociability and sales performance is crucial for creators. During the content creation process, creators should balance the congruence between their current and previous SVAs with the congruence between the video and the promoted product. Identifying these key influencing factors allows creators to design more persuasive and market-effective SVAs that optimize both user engagement and commercial outcomes.

Among various factors that shape the effectiveness of SVAs, such as video content, visual design, music, and storytelling, the video title plays a distinctive yet often underappreciated role. As the first textual cue users view when scrolling through videos, the title serves as a cognitive entry point, summarizing the core message, signaling thematic relevance, and helping users decide whether to engage further [14,15]. The title-related characteristics do not replace other multimodal features, such as narrative, music, or content design. It complements them by enhancing recognizability and trust, particularly under conditions of brief exposure and high content overload.

Existing research has concentrated on various consistent cues in short video contexts, such as creator-product congruence [3,16], creator-video congruence [17], and content-product congruence [18]. However, these studies focus on consumer perspective and rely on questionnaires or manual coding, with limited use of large-scale secondary data. To address this gap, this study utilizes real-world data collected from Douyin to explore how the SVA congruence characteristics (congruence with previous SVAs and congruence with the related product) affect product sales performance. By analyzing secondary data from platforms, this study enhances the objectivity and authenticity of research findings, thereby addressing the shortcomings of existing research in terms of data authenticity [3].

In addition, the sociability of SVAs, an essential indicator of content popularity, has been examined from multiple factors, including creator characteristics (number of followers, number of videos) and SVA characteristics (duration, posting days, title features, information disclosure, etc.) [19–21]. However, limited attention has been paid to how SVA congruence characteristics influence sociability. And how does sociability affect the relationship between SVA congruence and product sales? Prior research has acknowledged the positive effect of sociability on product sales [13], but they have overlooked the mediating role of sociability between SVA congruence and sales performance.

To fill these gaps, this study focuses on SVA creators as key stakeholders and explores how their consistent creation strategies—particularly title design—can enhance sociability and, in turn, sales performance. Drawing on the Signaling theory, we conceptualize the SVA title congruence as a video signal that influences sociability, which then drives product sales. By analyzing a large dataset from Douyin, we find that video-video title congruence (congruence with previous SVAs) positively affects sociability, while video-product title congruence (congruence with the related product) has a negative effect. In addition, SVA

sociability mediates the relationship between title congruence and sales performance. These relationships are further moderated by creator factors (creation frequency) and product characteristics (brand type). Additional analysis reveals that textual features of SVA titles (sentiment, first-person pronouns, and vividness) also significantly affect sales performance.

Our study makes the following contributions. First, this study concentrates on the SVAs sales performance from the perspective of creators, and we uncover the mediating role of sociability and enrich the literature on social media marketing. Second, our findings extend the relevant research on congruence, revealing the differential effects of congruence characteristics in different contexts through empirical analysis. Our findings also provide practical implications for both creators and platforms. This study provides suggestions for creators to optimize SVA creation strategies and boundary conditions to improve their performance. We also suggest platforms to encourage creators to post high-quality and attractive video content. Based on this strategy, platforms not only enhance consumers' preference and retention but also improve their profitability, ultimately realizing a win-win situation for both creators and platforms.

The remainder of this paper is organized as follows. Section 2 reviews the related literature on SVAs, congruence, and Signaling theory. Section 3 introduces research hypotheses. Section 4 describes the research methodology, including the data source and variable measurements. Section 5 exhibits empirical results, and Section 6 discusses results with theoretical and practical implications, as well as limitations and future directions of this study.

2. Literature Review

2.1. Short Video Advertisements (SVAs)

Digital advertising has been defined as the "Digital advertising involves paid media placements where advertisers bid for space to display promotional messages on digital platforms (e.g., search engines, social media, websites)" [22]. In digital contexts, this function extends to short video advertisements (SVAs), which integrate entertainment and commerce to engage consumers in interactive consumption journeys. SVAs, with durations ranging from a few seconds to minutes, contain embedded purchase links that enable consumers to complete instant transactions while watching the video [1,2,13]. Due to their cost-efficiency, rapid dissemination, and effective aggregation of viewership traffic, SVAs have been increasingly prevalent for product promotion and online sales [23].

With the popularity of this advertising model, scholars have explored how various SVA's characteristics—such as posting days, duration, thumbnail, and title—affect advertising performance [3,6,15,19,24] (See Table 1). In the context of short video consumption, the purchase process is as follows: Viewers are first exposed to SVAs in the platform, triggering attention and interest through multiple cues. Then, evaluation and trust form as viewers perceive creator credibility or sociability (likes, comments, and shares), ultimately leading to direct purchase via the embedded purchase link [2,13]. Unlike traditional advertising that involves prolonged consideration, SVA consumption involves instant decision-making, where video-related cues may play roles in converting attention into sales.

Table 1. Overview of literature on short video advertisements.

Research	Research Context	Theory	DV	IV	Moderator (mo) /Mediator (me)	Key Findings
Addo, Akpatsa [6]	JD, Tmall and Taobao	Signaling theory	Product sale	Short video ads (binary variable)	Price (mo)/product quality signal and consumer satisfaction (me)	The use of short video advertisements will promote product sales; the positive relationship is mediated by product quality signal and consumer satisfaction.
Dong, Liu [3]	Weibo short video	-	Consumer social engagement behavior (likes, comment, and share).	Video content features (content matching, information relevance, storytelling and emotionality)	Video release time (mo)	Short video content features significantly affect consumer engagement, compared to videos released in the afternoon, videos released in the morning increase the positive effects of warmth, excitement, and joy on consumer engagement.
Xiao, Li [24]	TikTok	Uses and gratifications theory, signaling theory	Consumer engagement	Performance expectancy; entertainment; tie strength; sales approach	Product type (mo)	Four key features of short videos (performance expectancy, entertainment, tie strength, and sales approach) significantly affect consumer engagement behavior, and product type moderates these relationships.
Meng, Kou [1]	TikTok	Grounded theory	Consumer purchase behavior	Trustworthiness, expertise, attractiveness, authenticity, and brand heritage	-	Trustworthiness, expertise, and attractiveness of video content can promote consumer purchase behavior, while authenticity and brand heritage have a U-shape effect on consumer purchase.
Yang, Wang [10]	Douyin	Heuristic-systematic model theory	Enterprise performance	Voice feature; music feature	-	When short video advertisements have a faster speech rate, high voice quality, faster music tempo, lower pitch, and lower loudness, enterprises achieve a more positive performance in social e-commerce.

Among these factors, the title of the SVA is a primary indicator of its content description and a critical determinant of viewership. The video title functions as a key cognitive signal that helps viewers rapidly assess content relevance, trustworthiness, and whether to browse or engage. Previous scholars have explored titles with positive sentiments and high arousal are more likely to attract viewers [14,15]. Meanwhile, concise and rhetorically rich titles enhance information transmission and viewer interests [21,25], while longer titles provide more information that enhances memorability and sociability [26]. These findings suggest that even when viewers do not watch full videos or are only briefly exposed, the title itself shapes viewers' perceptions, trust, and engagement, underscoring its importance as a research variable in the SVA context. However, existing studies have mainly focused on individual title characteristics, while relatively little attention has been paid to title congruence (congruence with previous SVAs titles and

congruence with the related product title) and how these congruence characteristics may affect sociability and sales performance. To address this gap, this study regards SVA title congruence as the primary research variable and investigates its effects on sociability and product sales performance.

Previous studies on SVAs have been grounded in multiple theoretical perspectives that offer complementary insights into users' cognitive, affective, and behavioral responses. Signaling theory explains how observable cues, such as the title, thumbnail, and hashtag, serve as signals that attract attention and reduce information asymmetry between content creators and viewers, thereby influencing perceptions of credibility and product quality [6]. The Uses and Gratification Theory emphasizes that users engage with short video content to satisfy intrinsic needs such as entertainment, information seeking, and social interaction, which sheds light on the role of sociability as a key outcome in the SVA context [24,27]. Moreover, the Grounded Theory and Heuristics-Systematic Model Theory are used to explain how consumers process and interpret persuasive advertising cues, such as titles, to understand and evaluate the video content [1,10]. Integrating these theoretical perspectives enables a comprehensive understanding of how the effect and mechanism of SVA title congruence influences sociability and sales performance. The present study adopts the Signaling Theory as the primary theoretical framework, given its strong explanatory power in elucidating how title-based cues shape consumer trust and behavioral responses in SVA-driven commerce.

2.2. Congruence

Congruence refers to the degree of similarity between two objects or activities [28]. Consistent information can strengthen associative impressions, further promoting cognitive and forming favorable attitudes [29,30]. In the traditional e-commerce context, congruence has been investigated across multiple dimensions, including congruence between product descriptions, pictures, and comments, which has been found to influence product evaluations and sales [31–34] (See Table 2).

Table 2. Overview of congruence literature in the context of non-short video advertisements.

Research	Research Context	Theory	DV	IV	Moderator (mo) /Mediator (me)	Key Findings
Eslami and Ghasemaghaei [31]	Amazon.com	Heuristic-systematic model	Product sale	Online review positiveness, review score inconsistency	Review score, review sentiment and review title sentiment (mo)	For a high-involvement product, review score incongruence negatively influences its sales, review textual sentiment positively influences its sales, and review title sentiment and review score influence low-involvement product sales.
Maier [34]	Online retail website	-	Attitudes towards the product assortment	In congruence of image characteristics on product overview	Disfluency perception of the product assortment	Product image incongruence in product overview negatively affects the assortment evaluations and choice satisfaction of the product.

Table 2. *Cont.*

Research	Research Context	Theory	DV	IV	Moderator (mo) /Mediator (me)	Key Findings
Park and Lin [33]	Taobao and Weibo	Match-up hypothesis	Purchase intention	Self-product fit, source-product fit, live content-product fit	Perceived trustworthiness (me), attractiveness (me), attitude toward content (me)	Source-product fit positively affects the source trustworthiness and attractiveness, live content-product fit positively affects attitude towards content, and source trustworthiness, hedonic attitude, and self-product fit enhance the purchase intention.
Rungruangjit [32]	Taobao	Source credibility, match-up hypothesis and parasocial relationship theory	Purchase intention	Celebrity-product congruence	Source credibility (me) and parasocial relationship (me)	A good match-up between celebrity and product can enhance perceived source credibility, and perceived expertise positively mediates the positive effect of celebrity-product congruence on purchase intention.

In the context of SVAs, scholars have investigated the congruence relationship, such as creators-products [16,18,30], creators-videos [17], videos-product information [3], and creators-consumers [35]. However, few studies have explored how congruence between different SVAs of the same creator influences sociability and sales performance. Additionally, when producing a new SVA, creators also consider and decide which title is consistent with the related product title. These two forms of congruence: video-vide title congruence and video-product title congruence, reflect distinct strategic choices that may signal the creator's image and product information to audiences.

To fill the above gaps, this study adopts the creator perspective, focusing on two congruence dimensions: (1) title congruence between the current and previous SVAs; (2) title congruence between the SVA and product. Building on the Signaling theory, this study investigates the effects of the above two title congruence on SVA sales performance, further explores the mediating role of sociability, as well as the heterogeneous effects of different characteristics of creators and products.

2.3. Signaling Theory

Signaling theory, originally proposed by Spence [36], explains how the sender conveys credible information to the receiver to reduce information asymmetry. In digital environments, where consumers cannot directly verify product quality or creator reliability, signals help shape perceptions and guide decisions [37,38]. In the short video context, signaling theory has been widely used to explain how various video-related characteristics, including video design, audio, quantity, or engagement metrics, influence consumer behaviors [6,24,39]. Among these signals, the title of the SVA, serving as a concise description of its content [14], also functions as a key cognitive signal. Prior studies have confirmed that an attractive or emotionally arousing title increases viewership and engagement [14,15,25].

Building on this perspective, our paper sets out to explore the effect of title congruence signals from video creators in SVAs. Specifically, it includes two aspects: (1) the title congruence between the creator's SVAs, and (2) the title congruence between the SVA and

the related product. These two features may help creators establish their own image and attract viewership on the platform. The former one signals reliability and stable content positioning, helping users recognize and recall a creator's brand style [40]. The latter one, while informative, may signal excessive commercial intent and potentially reduce user engagement [41].

Thus, from a signaling theory perspective, video creators should balance these cues to achieve both social resonance (sociability) and commercial effectiveness (sales performance). Drawing on signaling theory, we examine how the title congruence of SVAs influences sales performance and propose our research model and hypotheses in the following sections.

3. Hypotheses Development

3.1. Video-Video (V-V) Title Congruence and Sociability

Creators concentrate on the congruence between their current and previous SVAs to maintain a coherent identity and viewer expectations. Since the title of the SVA serves as a key descriptor of its content [14], research has confirmed that an attractive title increases viewership [25]. Video-video (V-V) title congruence refers to the consistent degree of similarity between the creator's current and previous SVAs. From a signaling perspective, repeated use of congruent V-V titles communicates reliability and stable content positioning, reducing cognitive uncertainty and reinforcing creator identity [40,42], as well as weakening competitor interference [40,43]. Prior research on advertising congruence shows that perceived coherence across messages enhances credibility, cognitive fluency, and use liking [44,45].

Moreover, processing fluency theory suggests that congruent stimuli are easier to process and tend to elicit more positive affective responses [46,47]. On short-video platforms where viewers make split-second judgments, such as fluency fosters recognition and emotional comfort, increasing social engagement behaviors (reflected as the SVA sociability, i.e., likes, comments, and shares). Even if viewers do not consciously read or compare titles, V-V title congruence may form implicit recognition and perceived familiarity as a peripheral signal [48,49]. High V-V title congruence enables viewers to quickly associate new SVA with prior positive experiences from the same creator, thereby fostering recognition, emotional comfort, and interaction tendency. Hence, we propose that:

Hypothesis 1: *V-V title congruence has a positive effect on SVA sociability.*

3.2. Video-Product (V-P) Title Congruence and Sociability

Video-product (V-P) title congruence refers to the degree of similarity between the title of the SVA and the promoted product's title. High V-P title congruence indicates that the SVA title explicitly conveys product-related information, potentially increasing informativeness and perceived credibility, thereby enhancing product sales [32,41]. However, the core of the SVA platform, such as Douyin, is entertainment rather than commerce [50]. Research shows that high information-product congruence can induce advertising fatigue and psychological resistance [1,51]. Hence, overtly commercial or "ad-like" titles (high V-P title congruence) can disrupt entertainment experience and lower social engagement, from a signaling perspective.

In contrast, low V-P title congruence, where the title maintains a storytelling or the overall style on the SVA platform, may reduce perceived intrusiveness and enhance entertainment value [52]. Such titles can subtly enhance familiarity and a positive effect over time, encouraging social responses without viewers' explicit awareness of the promotional intent. These indirect signals stimulate curiosity and pleasure [34,53,54], increasing

the sociability of the SVA through more viewer social engagement. Based on the above discussion, we propose the following hypothesis:

Hypothesis 2: *V-P title congruence has a negative effect on SVA sociability.*

3.3. Sociability and Sale Performance

Sociability refers to the degree to which users engage in social engagement, such as likes, comments, and shares, when watching SVAs. On short video platforms, sociability not only reflects the content's popularity but also indicates the perceived authenticity and resonance of both the creator and the promoted product [15,55,56]. From a signaling theory perspective, when the SVA receives numerous likes or comments, potential viewers interpret these cues as a signal of quality, credibility, and relevance [57]. Such bandwagon effect encourages subsequent viewers to engage with or purchase the advertising product, thereby amplifying its sales potential [15,55,56].

Creators who maintain high V-V title congruence strengthen stable and recognizable identity signals, helping viewers quickly identify familiar content and strengthen their parasocial relationships with the creator [13,40,42]. These consistent signals increase the likelihood of positive affective reactions, such as enjoyment and trust, which manifest as sociability behaviors.

Conversely, SVAs with low V-P title, which integrate product information subtly, tend to elicit curiosity and emotional resonance rather than resistance [50,52]. Such emotional engagement fosters deeper social interaction (i.e., likes, comments, shares), reinforcing viewer-creator connections and indirectly enhancing purchase potential.

Empirical evidence supports that social engagement indicators, likes, shares, and comments, serve as leading indicators of purchase conversion [13]. These metrics increase exposure and boost viewers' confidence in the product. Therefore, sociability bridges the relationship between SVA title congruence and sales performance. According to the above discussion, we propose that:

Hypothesis 3: *SVA sociability has a positive effect on sales performance.*

Hypothesis 4: *SVA sociability mediates the effects of title congruence on sales performance.*

3.4. Moderation Effect of Creation Frequency

Creation frequency, defined as the number of SVAs posted per day by the creator. It serves as a crucial behavioral signal that conveys the creator's activity level and shapes the degree of exposure and audience familiarity [58]. Higher creation frequency strengthens the creator's visibility and memorability in platform [59,60]. When combined with high V-V title congruence, frequent posting further reinforces creators' creation image and credibility image on the platforms [40], improving the visibility of the related product and amplifying sales performance. Hence, we propose that:

Hypothesis 5: *Creation frequency positively moderates the positive effect of V-V title congruence on SVA sales performance.*

When V-P title congruence is high, high creation frequency may lead to information overload and cognitive fatigue [61,62]. Repeated exposure to product-focused content heightens viewers' advertising resistance and reduces attention [1,51]. In contrast, frequent posting of low V-P title congruence (more entertaining) SVA enhances engagement and

purchase likelihood by sustaining curiosity and pleasure [53]. Hence, we propose the following hypothesis:

Hypothesis 6: *Creation frequency negatively moderates the negative effect of V-P title congruence on SVA sales performance.*

3.5. Moderation Effect of Product Brand

Product brand is a critical determinant of product sales [63,64] and serves as a key attribute in product promotion [3]. In online environments lacking direct product experience, brand information signal functions as a key quality signal [65], reducing perceived risk and strengthening the SVA content credibility [64]. Branded products enhance the persuasion and reinforce the positive effect of V-V title congruence on SVA sales performance. Hence, we propose the following hypothesis:

Hypothesis 7: *Product brand positively moderates the positive effect of V-V title congruence on SVA sales performance.*

Previous studies have shown that branded products have higher cognitive familiarity and trust [66], which may mitigate skepticism toward overt advertising [67]. In such cases, when a branded product is involved, even high congruence V-P titles are perceived as product-centered and recognize commercial intention; the negative effect of the SVA on sales performance will be weakened due to the product brand signal [64], thereby weakening its negative effect on SVA sales performance. Based on the above discussion, we propose that:

Hypothesis 8: *Product brand positively moderates the negative effect of V-P title congruence on SVA sales performance.*

Figure 1 illustrates the research model and the relationships between variables based on the hypotheses discussed.

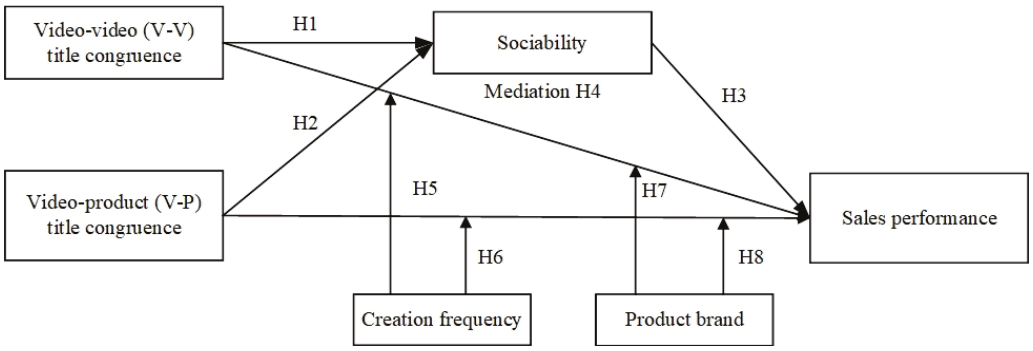


Figure 1. Research model.

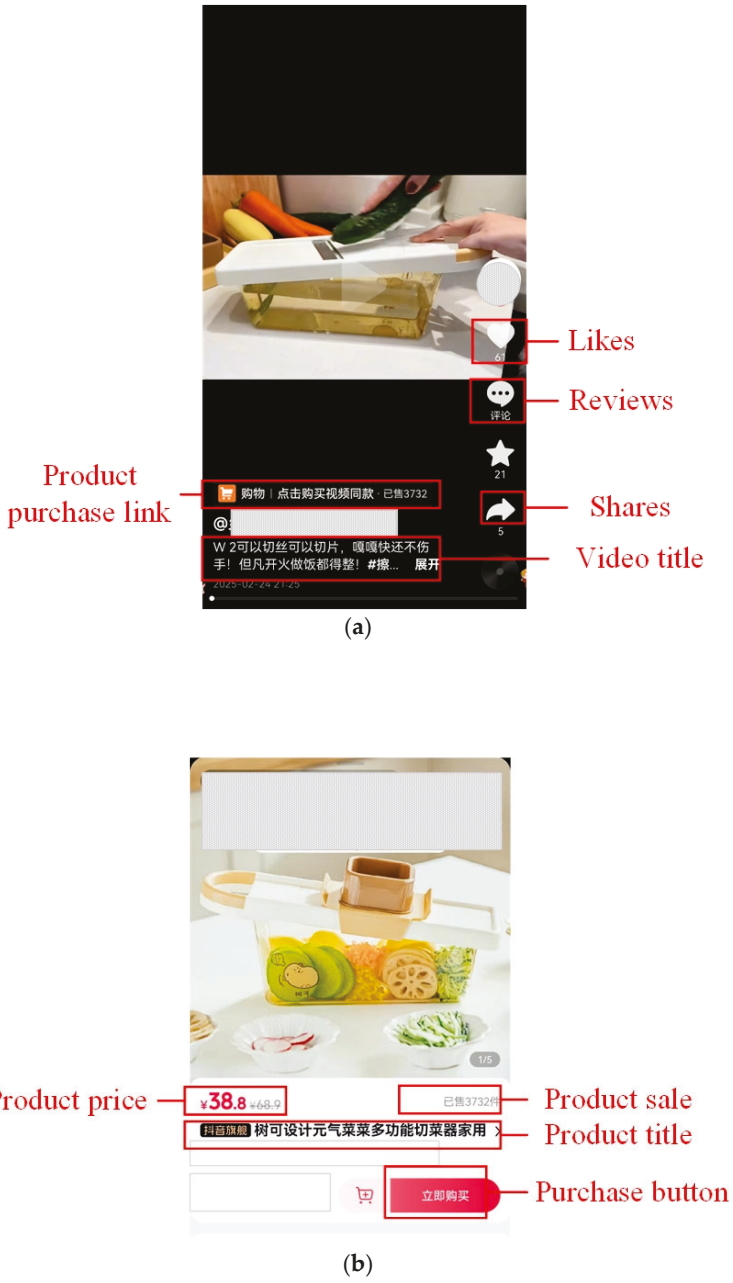
4. Research Methodology

4.1. Data Source

To test these hypotheses, we chose Douyin as our data source to explore the effects of SVAs title congruence on sales performance, for the following reasons. First, Douyin has over 600 million daily active users in China, nearly 300 million creators, and produces 20–30 million pieces of content daily [68], which is one of the largest and most successful SVA platforms globally [23,69]. Second, Douyin offers a diverse range of products in

SVAs [24], providing us with comprehensive and accurate data, making it an appropriate platform for our research.

Douyin provides an innovative business model by integrating short videos with e-commerce transactions [70]. As shown in Figure 2a, creators can embed product purchase links into their videos, allowing consumers to purchase products when watching the videos [69]. When they click the link, the interface will change into Figure 2b, where consumers can view product information, including price, title, and sale, as well as make purchase decisions by clicking the purchase button. Producers, platforms, and creators share the profits from commodity sales in a certain proportion [69,70]. This business model is the current dominant marketing model on the Douyin platform.



(**a**)

(**b**)

Figure 2. (**a**) The screenshot of a short video advertisement. (**b**). The screenshot of the purchase link embedded in a short video advertisement.

We randomly collected short video data from 400 creators with more than 100,000 fans who posted from January to September 2023 at three-month intervals. Consistent with prior literature [71,72], we adopted 100,000 followers as the threshold to define influencers, as this value is widely used to distinguish influencers from other creators. These creators are responsible for the majority of commercial content and account for a disproportionate share of platform-driven sales revenue [73–75]. While smaller creators contribute to platform activity, their commercial influence and advertising engagement rate are lower, which makes them less suitable for studying sales-driving mechanisms.

We randomly collected data using a stratified sampling approach based on content category distribution to ensure representativeness across different sectors. Considering the effect of the active period of short video on data, we removed video data that were posted more than 30 days before the data collection date to minimize data collection bias and keep variable validity, according to prior research [1,21,24]. In addition, since our study focuses on short video ads that provide product purchasing links, we removed those videos without the product purchase link. After data cleaning, 80 creators were excluded due to incomplete or inconsistent data records, or missing product information. The final data on a total of 57,192 short videos with product purchase links posted by 320 creators covering 19 sectors were collected. Each data contains characteristics of the short video ad (i.e., title, posting time, duration, likes, and comments number), related product characteristics (i.e., title, price, brand, and sales), and creator characteristics (i.e., fans number, sector, and experience).

4.2. Variable Measurements

Sales performance. We operationalized SVA success through direct sales performance, representing the ultimate commercial effectiveness. Specifically, we used product sales on the data collection day as the measurement of SVA sales performance, which is recorded by the platform.

Video-video (V-V) title congruence. We measured V-V title congruence as the level of congruence between the creator's current short video advertisement and previous advertisements posted within the preceding 14-day period. SVAs posted from the 14 days are chosen because the active period of SVAs is relatively short [21]; and to observe whether creators can maintain a consistent SVA creation style, we selected 14 days as the observation period (To test the robustness of research results, we presented the results of three alternative time windows (7 days, 21 days, and 28 days) in the robustness test. These results for different time windows are generally consistent with the main results.). We employed word embedding techniques to measure title congruence among multiple video titles based on semantic vectors. Specifically, we first used Jieba, a widely adopted Chinese text segmentation tool, to segment each video title into word phrases while removing common stop words during this process. Subsequently, we applied word2vec technology to generate 64-dimensional word vectors for the segmented phrases from video titles, thereby capturing semantic information. Considering that each phrase carries different importance across various contexts, we calculated TF-IDF (Term Frequency-Inverse Document Frequency) values for each phrase within video titles as weights representing their significance. Each video title was then represented using the weighted sum of word vectors from its constituent phrases, to simultaneously consider the semantic information of each word and its importance in the current video title [76,77]. Finally, we evaluated the congruence between paired titles using cosine similarity (ranging from -1 to 1), a metric that quantifies the directional similarity between two vectors [78]. The value of V-V title congruence approaching 1 indicates higher similarity between the current and previous SVAs, thus reflecting greater congruence and higher creation

image. We calculated the congruence between the current video title and each video title published within the past 14 days, then computed the average value as the final V-V title congruence measurement.

Video-product (V-P) title congruence. Using the same congruence calculation method, we represented the product title as semantic vectors and measured V-P title congruence by calculating the cosine similarity between the SVA title and the product title. The closer the value is to 1, the more product-promoted information is directly conveyed in the SVA.

Sociability. The mediator variable in this study refers to SVA sociability. Following prior research [13,79], we used the total sum of likes, comments, and shares of SVA to reflect its sociability. Using the sum of these three dimensions as the measurement can comprehensively assess the effectiveness of SVA social performance [57]. This variable not only reflects the viewer engagement but also serves as a behavioral process linking through how SVA title congruence translates into commercial outcomes, as suggested by the research conceptual model.

Creation frequency. We calculated the number of SVAs posted by the creator in a day to capture the creation frequency. The higher the value, the higher the creation frequency of the creator.

Product brand. We used the dummy variable to indicate whether the product is a branded product. If the product contains brand information, the value of the variable is 1; otherwise, it is 0.

Control variables. According to prior studies, we controlled potential influences on product sales in terms of SVAs, products, and creators. In terms of SVAs, we included video day of week (whether the video is posted on the working day); video time of day (whether the video is posted in the evening); video length (minutes based); video age (the days intervals between video posting date and data collection date); title length and tag length [3,21,25]. We also added creator-related characteristics to control their effects, such as fans, expertise (the day intervals of creators have been on the platform), and sex [18,20]. Product price is controlled for its potential effect on product sales [1]. In addition, we also controlled the time-fixed effect of short video ads posting month, the creator-level fixed effect, and the creator's sector-level fixed effect to improve the robustness of our research [3,25].

Table 3 states the descriptive statistics of research variables. Taking high skewness, kurtosis, and non-normal distribution into consideration, we applied the logarithmic transformation of these variables (except for title congruence, brand, video day of week, video time of day, video length, and creator sex) [80]. Before conducting the regression model, we analyzed the covariance and correlation among all variables. As Table 4 shows, the variance inflation factors (VIF) value between research variables is less than 10, and the correlations of all variables were all less than the threshold value (See Table 5), providing evidence that multicollinearity was not a concern in our research model.

Table 3. Descriptive statistics of research variables.

Variable	Mean	Std. Dev.	Min	Max
Sales	315,185.900	1,156,790	0	51,100,000
V-V title congruence	0.605	0.295	−0.277	1
V-P title congruence	0.660	0.223	−0.408	1
Sociability	695.997	7332.278	0	1,186,097
Creation frequency	6.136	4.278	1	50
Brand	0.568	0.495	0	1
Video day of week	0.728	0.445	0	1
Video time of day	0.292	0.455	0	1
Video length	0.559	0.588	0	14.280
Video age	46.983	27.163	0	97
Video title length	42.639	20.119	0	462
Video tag number	2.556	1.574	0	10
Product price	80.952	363.585	0.060	22,888
Creator fans	1,001,553	1,476,317	100,000	52,200,000
Creator expertise	112.941	85.377	2	316
Creator sex	0.597	0.490	0	1

Table 4. VIF values of research variables.

Variable	VIF	1/VIF	Variable	VIF	1/VIF
V-V title congruence	1.230	0.810	Video age	1.010	0.994
V-P title congruence	1.440	0.693	Video title length	1.310	0.764
Sociability	1.860	0.537	Video tag number	1.360	0.737
Creation frequency	2.160	0.464	Product price	1.090	0.918
Brand	1.050	0.952	Creator fans	1.230	0.814
Video day of week	1.000	0.996	Creator expertise	2.020	0.494
Video time of day	1.010	0.987	Creator sex	1.170	0.855
Video length	1.330	0.752	Mean VIF	1.350	

Table 5. Correlations among research variables (N = 57,192).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) Sales	1															
(2) V-V title congruence	0.049 ***	1														
(3) V-P title congruence	0.076 ***	0.113 ***	1													
(4) Sociability	0.030 ***	−0.267 ***	−0.235 ***	1												
(5) Creation frequency	−0.062 ***	0.049 ***	−0.041 ***	−0.417 ***	1											
(6) Brand	0.326 ***	0.057 ***	0.050 ***	0.058 ***	−0.120 ***	1										
(7) Video day of week	0.010 **	0.012 ***	0.004	−0.014 ***	0.028 ***	0.004	1									
(8) Video time of day	0.057 ***	0.016 ***	0.005	0.024 ***	−0.030 ***	0.038 ***	−0.001	1								
(9) Video length	0.006	−0.282 ***	−0.080 ***	0.427 ***	−0.198 ***	0.023 ***	0	−0.050 ***	1							
(10) Video age	0.069 ***	−0.007 *	−0.021 ***	0.029 ***	−0.008 **	0.008 **	0.043 ***	0.001	−0.016 ***	1						
(11) Video title length	0.092 ***	0.316 ***	0.118 ***	−0.005	−0.074 ***	0.061 ***	0.004	−0.032 ***	−0.018 ***	−0.020 ***	1					
(12) Video tag number	0.031 ***	0.358 ***	0.131 ***	−0.077 ***	−0.032 ***	0.069 ***	−0.001	−0.003	−0.050 ***	0.011 ***	0.428 ***	1				
(13) Product price	−0.417 ***	−0.147 ***	−0.116 ***	0.102 ***	0.093 ***	−0.146 ***	−0.005	−0.032 ***	0.129 ***	−0.024 ***	−0.064 ***	−0.075 ***	1			
(14) Creator fans	−0.049 ***	0.073 ***	0.047 ***	0.298 ***	−0.037 ***	0.034 ***	0.001	0.041 ***	0.026 ***	−0.001	−0.034 ***	−0.074 ***	0.076 ***	1		
(15) Creator expertise	0.015 ***	0.171 ***	0.045 ***	−0.390 ***	0.682 ***	−0.083 ***	0.005	−0.008 *	−0.251 ***	−0.003	0.035 ***	0.075 ***	−0.001	−0.071 ***	1	
(16) Creator sex	−0.101 ***	0.139 ***	−0.246 ***	−0.007	0.014 ***	−0.073 ***	0.002	−0.043 ***	−0.134 ***	0.001	0.033 ***	−0.079 ***	−0.029 ***	0.110 ***	−0.039 ***	1

Standard errors in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5. Model Specific and Research Results

5.1. Empirical Results

5.1.1. Main Result

To test these relationships in our research model, we applied the three steps proposed by Baron and Kenny [81] to test our hypotheses. The empirical formulations are as follows:

$$\begin{aligned} Sales = \quad & \alpha + \beta_1 V - V\ title\ congruence + controls \\ & + \eta + v + \lambda + \varepsilon \end{aligned} \tag{1}$$

$$\begin{aligned} Sociability = \quad & \alpha_1 + \beta_2 V - V\ title\ congruence + controls \\ & + \eta_1 + v_1 + \lambda_1 + \varepsilon_1 \end{aligned} \tag{2}$$

$$\begin{aligned} Sales = \quad & \alpha_2 + \beta_3 V - V\ title\ congruence + \beta_4 Sociability + controls \\ & + \eta_2 + v_2 + \lambda_2 + \varepsilon_2 \end{aligned} \tag{3}$$

where *controls* captures variables influencing SVA sale performance, such as video day of week, video time of day, video length, video age, video title length, video tag number, product price, creator fans, creator expertise, and creator sex. In addition, η captures the time-fixed effect, v represents the creator-level fixed effect, λ represents the creator's sector-level fixed effect, ε denotes the idiosyncratic error, and α refers to the constant term.

Table 6 shows empirical results, Model 2 verifies that V-V title congruence positively and significantly affects SVA sociability (β_2 = 0.381, p < 0.01), which supports H1.

Table 6. OLS regression model results.

	Model 1 Sale	Model 2 Sociability	Model 3 Sale	Model 4 Sale	Model 5 Sociability	Model 6 Sale
V-V congruence	0.356 ***	0.381 ***	0.248 ***			
	(0.056)	(0.022)	(0.056)			
V-P congruence				−0.201 ***	−0.206 ***	−0.141 **
				(0.069)	(0.028)	(0.069)
Sociability			0.281 ***			0.284 ***
			(0.011)			(0.011)
Video day of week	0.035	−0.022 **	0.042 *	0.035	−0.022 **	0.042 *
	(0.023)	(0.009)	(0.022)	(0.023)	(0.009)	(0.022)
Video time of day	−0.030	0.028 ***	−0.038	−0.031	0.026 ***	−0.039
	(0.024)	(0.010)	(0.024)	(0.024)	(0.010)	(0.024)
Video length	0.706 ***	0.804 ***	0.476 ***	0.709 ***	0.807 ***	0.475 ***
	(0.064)	(0.030)	(0.064)	(0.064)	(0.030)	(0.064)
Video age	0.228 ***	0.086 ***	0.204 ***	0.229 ***	0.087 ***	0.204 ***
	(0.017)	(0.006)	(0.017)	(0.017)	(0.006)	(0.017)
Video title length	0.090 ***	0.026 **	0.082 ***	0.123 ***	0.061 ***	0.105 ***
	(0.027)	(0.012)	(0.027)	(0.027)	(0.012)	(0.027)
Video tag number	0.061 *	0.064 ***	0.044	0.073 **	0.076 ***	0.052
	(0.036)	(0.015)	(0.036)	(0.036)	(0.015)	(0.036)
Product price	−0.986 ***		−0.970 ***	−0.988 ***		−0.971 ***
	(0.020)		(0.020)	(0.020)		(0.020)
Creator fans	0.282	−0.091	0.302	0.304	−0.070	0.317
	(0.195)	(0.090)	(0.194)	(0.196)	(0.091)	(0.195)
Creator expertise	−0.171 ***	−0.199 ***	−0.114 ***	−0.183 ***	−0.212 ***	−0.122 ***
	(0.026)	(0.011)	(0.026)	(0.026)	(0.011)	(0.026)
Creator sex	1.995 ***	5.211 ***	0.569	1.710 ***	4.913 ***	0.355
	(0.442)	(0.171)	(0.443)	(0.442)	(0.169)	(0.443)
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Creator fixed effect	Yes	Yes	Yes	Yes	Yes	Yes

Table 6. *Cont.*

	Model 1 Sale	Model 2 Sociability	Model 3 Sale	Model 4 Sale	Model 5 Sociability	Model 6 Sale
Sector fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Constant	6.635 ***	3.486 ***	5.644 **	6.612 ***	3.491 ***	5.614 **
	(2.522)	(1.162)	(2.503)	(2.530)	(1.164)	(2.510)
N	57,192	57,192	57,192	57,192	57,192	57,192
R^2	0.549	0.745	0.555	0.549	0.744	0.555
adj. R^2	0.546	0.744	0.552	0.546	0.742	0.552

Standard errors in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

We then applied the three steps proposed by Baron and Kenny [81] to test hypothesis 2, the empirical formulations are as follows:

$$\begin{aligned} Sales = \quad & \alpha_3 + \beta_5 V - P\ title\ congruence + controls \\ & + \eta_3 + v_3 + \lambda_3 + \varepsilon_3 \end{aligned} \tag{4}$$

$$\begin{aligned} Sociability = \quad & \alpha_4 + \beta_6 V - P\ title\ congruence + controls \\ & + \eta_4 + v_4 + \lambda_4 + \varepsilon_4 \end{aligned} \tag{5}$$

$$\begin{aligned} Sales = \quad & \alpha_5 + \beta_7 V - P\ title\ congruence + \beta_8 Sociability + controls \\ & + \eta_5 + v_5 + \lambda_5 + \varepsilon_5 \end{aligned} \tag{6}$$

Model 5 displays that V-P title congruence has a negative and significant effect on social engagement behavior ($\beta_6 = -0.206$, $p < 0.01$), thus supporting H2.

Model 3 ($\beta_4 = 0.281$, $p < 0.01$) and Model 6 ($\beta_8 = 0.284$, $p < 0.01$) show that social engagement behavior positively and significantly affects purchase behavior, which supports H3.

In addition, V-V title congruence positively and significantly affects SVA sociability ($\beta_2 = 0.381$, $p < 0.01$) and sociability positively and significantly affects sale performance ($\beta_4 = 0.281$, $p < 0.01$), which indicates that sociability plays a mediating role in the relationship between V-V title congruence and sale performance (indirect effect ($\beta_2 \times \beta_4$) = 0.107, $p < 0.01$). Meanwhile, V-P title congruence negatively and significantly affects sociability ($\beta_6 = -0.206$, $p < 0.01$), while sociability positively and significantly affects sale performance ($\beta_8 = 0.284$, $p < 0.01$), which means that sociability plays a mediating role in the relationship between V-P title congruence and sale performance (indirect effect $\beta_6 \times \beta_8 = -0.059$, $p < 0.01$). Thus, H4 is supported.

Although the large sample size increases statistical power, the standardized coefficients confirm that SVA title congruence has substantial real-world effects. Specifically, a one standard deviation increase in V-V title congruence is associated with a 0.381 SD increase in sociability and a 0.356 SD increase in sales. Meanwhile, a one standard deviation increase in V-P title congruence is associated with a −0.206 SD increase in sociability and a −0.201 increase in sales. In addition, sociability strongly predicts sales ($\beta_4 = 0.281$; $\beta_8 = 0.284$, $p < 0.01$).

These results indicate that the V-V title demonstrates the strongest and most consistent influence on sales performance. It not only directly increases purchase potential but also indirectly drives sales through enhanced sociability. Conversely, V-P tile congruence exerts a negative effect on sales, indicating that over-commercialized titles may reduce viewer engagement. While all hypotheses were supported, the standardized coefficients reveal that maintaining consistent title creation patterns is the most critical factor for creators maximizing the commercial effectiveness of SVAs.

5.1.2. Moderation Effect

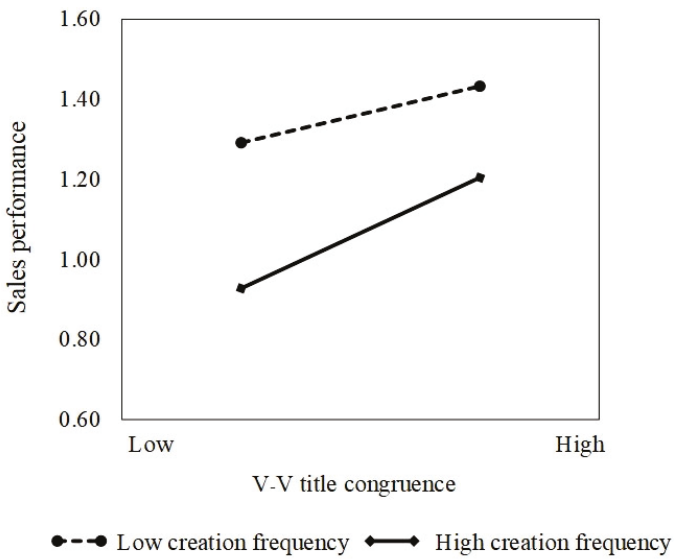
Based on the above analysis, we further developed the heterogeneity analysis to verify the moderating effect of creation frequency and product brand on the relationship between title congruence and sales performance. The specific equations are as follows:

$$\begin{aligned} Sales = \alpha_6 + \beta_9 V- \quad & V\ title\ congruence \\ & +\beta_{10} V - P\ title\ congruence \\ & +\beta_{11} Creation\ frequency + \beta_{12} Product\ brand \\ & +\beta_{13} Creation\ frequency \times V - V\ title\ congruence \\ & +\beta_{14} Creation\ frequency \times V - P\ title\ congruence \\ & +\beta_{15} Product\ brand \times V - V\ title\ congruence \\ & +\beta_{16} Product\ brand \times V - P\ title\ congruence \\ & +controls + \eta_6 + v_6 + \lambda_6 + \varepsilon_6 \end{aligned} \tag{7}$$

where $Creation\ frequency \times V - V\ title\ congruence$ represents the interaction term between creation frequency and V-V title congruence, and β_{13} represents the moderating effect of creation frequency on the effect of V-V title congruence on sale performance; $Creation\ frequency \times V - P\ title\ congruence$ is the interaction term between creation frequency and V-P title congruence, and β_{14} represents the moderating effect of creation frequency on the effect of V-P title congruence on sale performance.

In addition, $Product\ brand \times V - V\ title\ congruence$ is the interaction term between product brand and V-V title congruence, $Product\ brand \times V - P\ title\ congruence$ is the interaction term between product brand and V-P title congruence. β_{15} and β_{16} denotes the moderating effect of product brand on the effect of V-V title congruence and V-P title congruence on sales performance, respectively.

As the results of Model 7 show in Table 7, creation frequency positively moderates the effect of V-V title congruence on product sales (β_{13} = 0.148, $p < 0.05$). As shown in Figure 3a, the positive effect of V-V title congruence on SVA sales becomes stronger as the creation frequency increases, thus, H5 is supported. In addition, creation frequency negatively moderates the effect of V-P title congruence on product sales ($\beta_{14} = -0.501$, $p < 0.01$), Figure 3b shows that the negative effect of V-P title congruence on sale performance is stronger as the creation frequency increase, which support H6.



(**a**)

Figure 3. *Cont.*

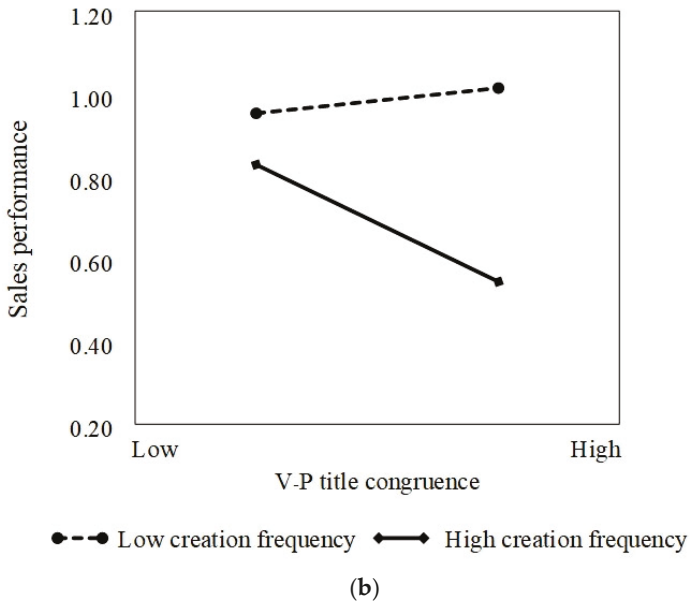


Figure 3. (**a**) Moderating effect of creation frequency on the effect of V-V title congruence. (**b**) Moderating effect of creation frequency on the effect of V-P title congruence.

Table 7. Research results of moderation effects.

Sale	Model 7	Model 8	Model 9	Model 10	Model 11
V-V title congruence	0.354 ***		−0.082		−0.064
	(0.057)		(0.071)		(0.071)
V-P title congruence		−0.250 ***		−0.821 ***	−0.738 ***
		(0.069)		(0.094)	(0.094)
Creation frequency	−0.281 ***	0.139 ***			0.035
	(0.046)	(0.054)			(0.065)
Creation frequency × V-V title congruence	0.148 **				0.153 ***
	(0.060)				(0.058)
Creation frequency × V-P title congruence		−0.501 ***			−0.458 ***
		(0.072)			(0.070)
Product brand			1.330 ***	0.982 ***	0.740 ***
			(0.055)	(0.083)	(0.092)
Product brand × V-V title congruence			0.702 ***		0.653 ***
			(0.082)		(0.083)
Product brand × V-P title congruence				1.092 ***	0.908 ***
				(0.115)	(0.115)
Month fixed effect	Yes	Yes	Yes	Yes	Yes
Creator fixed effect	Yes	Yes	Yes	Yes	Yes
Sector fixed effect	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Constant	5.139 **	3.503	11.536 ***	11.365 ***	8.848 ***
	(2.540)	(2.556)	(2.443)	(2.440)	(2.455)
N	57,192	57,192	57,192	57,192	57,192
R^2	0.550	0.550	0.578	0.578	0.580
adj. R^2	0.547	0.547	0.576	0.576	0.577

Standard errors in parentheses: ** $p < 0.05$, *** $p < 0.01$.

Model 9 shows that product brand positively moderates the effect of V-V title congruence on product sales (β_{15} = 0.702, p < 0.01). In Figure 4a, compared to unbranded products, V-V title congruence has a stronger positive effect on sales performance when promoting branded products. Model 10 verifies that product brand positively moderates the negative effect of V-P title congruence on product sales (β_{16} = 1.092, p < 0.01). As shown in Figure 4b, when promoting branded products in SVAs, the negative effect of V-P title congruence on sales performance is weaker compared to unbranded products. Thus, H7 and H8 are supported.

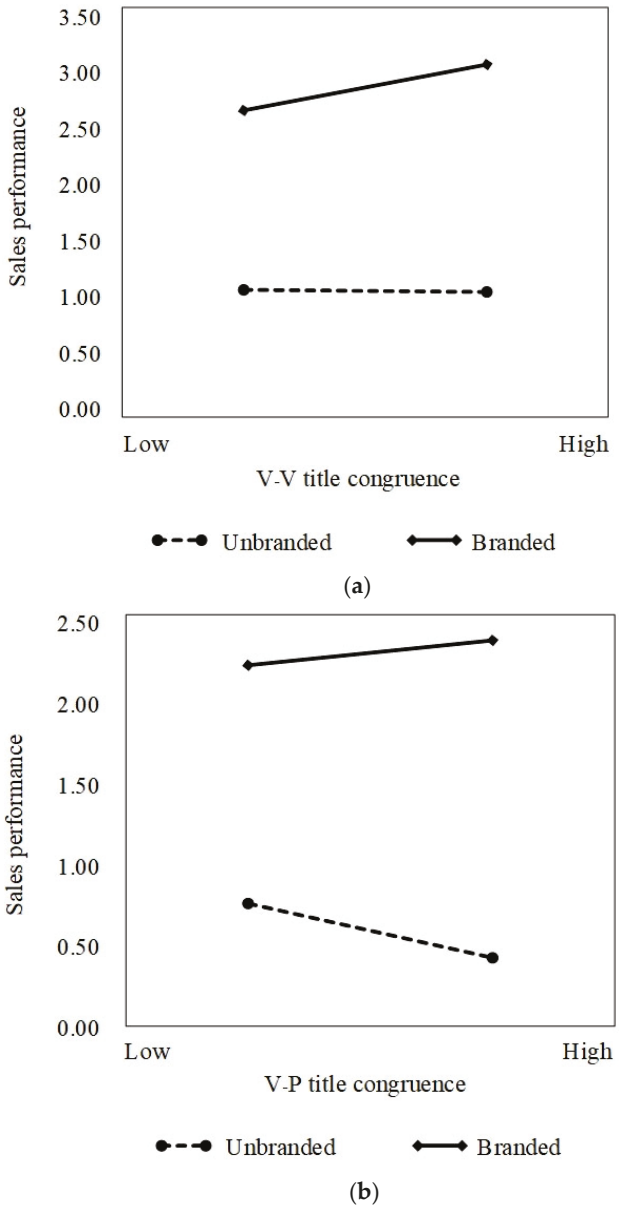


Figure 4. (**a**) Moderating effect of product brand on the effect of V-V title congruence. (**b**) Moderating effect of creation frequency on the effect of V-P title congruence.

In summary, this study tested eight hypotheses regarding the effects of SVA title congruence on sales performance through the mediating role of sociability, as well as the moderating role of creation frequency and product brand. Table 8 presents the summary of all hypothesis test results.

Table 8. Results of hypotheses testing.

	Hypothesis	Result
Hypothesis 1	V-V title congruence has a positive effect on SVA sociability.	Supported
Hypothesis 2	V-P title congruence has a negative effect on SVA sociability.	Supported
Hypothesis 3	SVA sociability has a positive effect on sales performance.	Supported
Hypothesis 4	SVA sociability mediates the effects of title congruence on sales performance.	Supported
Hypothesis 5	Creation frequency positively moderates the positive effect of V-V title congruence on SVA sales performance.	Supported
Hypothesis 6	Creation frequency negatively moderates the negative effect of V-P title congruence on SVA sales performance.	Supported
Hypothesis 7	Product brand positively moderates the positive effect of V-V title congruence on SVA sales performance.	Supported
Hypothesis 8	Product brand positively moderates the negative effect of V-P title congruence on SVA sales performance.	Supported

5.2. Robust Tests

In this section, we performed three robust checks of our research model. First, we applied the Heckman two-stage test to control for the potential self-selection bias based on observed variables. Second, we used the alternative measurement of V-V title congruence to test our research model. Finally, we used the alternative measurement of sociability. These test results that support our findings are robust.

5.2.1. Heckman Two-Stage Test

Self-bias occurs if creators differ systematically from content creation strategies. In our study, the decision of the creator whether to post the SVA or not is endogenous. Because the creation decision simultaneously affects the creator's income and future creation strategies. In this case, the Heckman two-stage test is the effective method to correct the bias according to prior research [82,83].

The first step is to estimate the probability of an anchor posting the short video ad by using the probit model. We choose the creator's characteristics (fan number, expertise) and video ad characteristics (viewership, likes, comments, share, duration, posting days, day of week, time of day, title length, and tag length) as the variables (See Table A1 in Appendix A) that may affect a creator's propensity to post SVAs.

The second step is to add the inverse Mills ratio (IMR) calculated by the probit model in the observation model to obtain unbiased results. Table A2 shows that the observation model results (shown in Appendix A) are consistent with our results and support the robustness of our research model.

5.2.2. Alternative Measurement of V-V Title Congruence

We also used the alternative measurement of V-V title congruence to check the robustness of the main results. We expanded the day intervals of V-V title congruence between the current and previous SVAs from the original 14 days posted by the creators to the last 7, 21, and 28 days. The results are shown in Table A3 (Appendix B), which are consistent with the main findings.

5.2.3. Alternative Measurement of Sociability

Referring to prior research [20,84], we applied the alternative measurement of sociability, using the total sum of likes, comments and shares divided by the number of the creator followers, we obtained a relatively standardized metric that eliminates the influence of followers' number [20]. We used the methods proposed by Baron and Kenny [81] to re-estimate the research model with the sociability ratio. The results of Table A4 are consistent

with those of the original model (shown in Appendix C). Thus, our findings are robust across different measurements of sociability.

5.3. Additional Analysis

To extend our primary findings, we conducted a supplementary analysis to investigate how textual features of video titles affect product sales, aiming to develop strategic recommendations for content creators. We categorized video titles into two distinct types: product-independent and product-related.

We initially randomly selected 500 titles for manual classification. Using Unsloth [85], we fine-tuned the Qwen2.5-7B-Instruct model (one of the most renowned open-source LLMs) based on these manually classified results. The fine-tuned model was then employed to classify all video titles. To evaluate the model's performance, we randomly selected another 500 titles for manual classification (coded by two human coders) as a test set, and the model achieved an accuracy of 92.85%. In addition, we also calculated the inter-rater reliability (IRR) metrics to evaluate the agreement between human coders and LLM according to prior research [86]. Following Chew et al.'s [87] equations, we calculated the IRR between manual classification and LLM and the result is 93.08%, which meets our experimental requirements.

Table 9 presents examples of the two types of short video titles. Product-independent short video titles do not contain information about the product-related features; while the product-related short video titles directly contain information about the product price, features, functions, and so on. These differences in textual features may lead to different marketing effects. Therefore, we extracted three textual features of the title, including sentiment, first-person pronoun, and vividness, to explore their effects on product sales (The measurement process of sentiment, first-person pronoun, and vividness is stated in Appendix D). We applied the *t*-test method to compare the differences between the above three textual features of the two short video types. Table 10 indicates the significant differences between the two types of short video titles on the above features.

Table 9. Examples of two short video types.

Short Video Type	Example
Type 1 product-independent	The days don't panic; we grow old together. Feel free to come to Sister Zhong's live streaming tonight for peanuts.
Type 2 product-related	9.9 RMB to make a photo frame for yourself with just a cell phone photo ~ love it so much! The days are getting cooler, and wearing shoes is good for sweaty feet. Foot odors dare not take off shoes, must try this shoe and sock freshness spray. Spray it to freshen up all day long #shoe and sock deodorizing spray.

Table 10. T-test results of video title characteristics.

Variable	Type 1 (N = 4947)	Type 2 (N = 52,245)	Mean Difference
Title sentiment	0.029	0.022	0.007 ***
Title first-person	0.007	0.003	0.003 ***
Title vividness	0.522	0.847	−0.326 ***

Standard errors in parentheses: *** $p < 0.01$.

On this basis, we further discussed the effects of these textual features on product sales. As shown in the result of Model 12 in Table 11, title sentiment negatively affects product sales ($\beta_{sentiment} = -1.067$, $p < 0.01$). Compared to positive titles, negative titles indicate SVAs contain interesting and curious content and are more likely to obtain more

attention [88]. And this relationship is positively moderated by video type ($\beta_{type \times sentiment}$ = 0.678, $p < 0.1$). This result indicates that the negative effect of title sentiment is more significant in product-independent SVAs.

Table 11. Results of effects of video title textual characteristics.

Sales	Model 12	Model 13	Model 14
Title sentiment	−1.067 ***		
	(0.328)		
Type	−0.110 **	−0.128 ***	−0.141 ***
	(0.046)	(0.045)	(0.052)
Type × Title sentiment	0.678 *		
	(0.395)		
Title first-person		−4.335 ***	
		(1.587)	
Type × Title first-person		8.515 ***	
		(1.931)	
Title vividness			−0.034
			(0.046)
Type × Title vividness			0.111 **
			(0.046)
Month fixed effect	Yes	Yes	Yes
Creator fixed effect	Yes	Yes	Yes
Sector fixed effect	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Cons	7.328 ***	7.156 ***	6.781 ***
	(2.520)	(2.518)	(2.517)
N	57,192	57,192	57,192
R^2	0.549	0.549	0.549
adj. R^2	0.546	0.546	0.546

Standard errors in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The results of Model 13 indicate that first-person pronouns negatively affect product sales ($\beta_{first-person} = -4.355$, $p < 0.01$). While the first-person pronouns in short video ad titles contribute to emotional connection and empathy senses, this narrative perspective is self-centered [89], and the usage of first-person narrative perspectives tends to be perceived as subjective, which is less credible, and likely to exaggerate and distorts the facts for their gain [90,91]. Video type positively moderates the negative relationship ($\beta_{type \times first-person}$ = 8.515, $p < 0.01$), which indicates that the negative effect of first-person pronouns is weakened in those product-dependent SVAs.

Moreover, the results of Model 14 show that vividness negatively influences product sales ($\beta_{vividness} = -0.034$, $p > 0.1$), and the negative relationship is moderated by video type ($\beta_{type \times vividness}$ = 0.111, $p < 0.05$). These findings suggest that titles containing excessive suspense or storyline may increase the information uncertainty and cognitive cost of consumers, and they may consider these as inappropriate marketing strategies [21].

6. Discussion

6.1. Key Findings

How to effectively design short video advertisements to promote product sales is the main question for creators. Based on the Signaling theoretical framework, this study explores the effects of title congruence characteristics of SVA on sociability and sales performance, as well as the heterogeneous effects from the perspectives of creators and product characteristics. Our findings provide insights into content creation strategies in the short video context.

First, this study found that V-V title congruence positively affects SVA sociability [1]. This suggests that consistent SVA content helps creators develop their creation image [92]. Meanwhile, it helps creators differentiate themselves from others on the platform, reducing interference from others [40,43], fostering familiarity and trust, and thereby increasing interaction willingness. For creators, maintaining a certain level of V-V title congruence can help them establish a personal creation image on the platform, enhance their distinctiveness, and ultimately improve SVA sales performance.

Second, unlike the former, this study found that V-P title congruence negatively affects SVA sociability. This conclusion differs from previous studies on product consistency, which may be due to differences in the attributes of the research platforms. Previous studies were primarily conducted on online e-commerce platforms such as Taobao and Amazon, which are e-commerce platforms designed to facilitate product purchases for consumers. These studies found that cues related to product congruence help foster content trust, which in turn influences product sales [41,93]. However, on SVA platforms centered on social entertainment [50], SVAs with high V-P title congruence highlight product attributes in their contents, making it easier to identify direct product promotion, triggering cognitive resistance, and leading to advertising avoidance [52]. Therefore, for creators on SVA platforms, it is advisable to avoid presenting product information in the SVA titles and instead adopt indirect or more entertaining title strategies to reduce resistance and avoidance of SVAs.

Third, this study found that sociability serves as a partial mediator between title congruence and sales performance, indicating that part of the effect of title congruence on sales remains direct. The sociability of SVAs positively influences SVA sales performance, which is consistent with previous research findings [13]. Sociability reflects trust and popularity of content, which further enhances the promoted product's visibility and purchase likelihood [94]. This finding emphasizes the mediating role of SVA sociability in sales performance. Moreover, it suggests that some viewers may directly purchase products without performing any social interactions, enhancing sociability through interaction, trust, and ultimately convert into buyers. Therefore, creators could design SVA titles that capture audience attention and encourage interaction, as strengthening sociability remains a practically effective way to achieve the goal of product sales on SVA platforms.

Fourth, our results show that creation frequency positively moderates the relationship between V-V title congruence and product sales. When creators consistently create and publish SVAs with consistent styles, it enhances personal image perception and improves SVA sales performance. Creation frequency negatively moderates the effect of V-P title congruence on sales performance. That is, when creators frequently publish product-oriented high-congruence SVAs (high V-P title congruence), this may exacerbate resistance and avoidance caused by direct advertisements, thereby strengthening the negative effect of V-P title congruence. Additionally, this study found that the product brand significantly moderates the effects of V-V and V-P title congruence on sales performance. Product brands can enhance content trust [66], thereby weakening the advertisement avoidance triggered by high V-P title congruence and strengthening the creator's personal image formed by V-V title congruence.

Finally, through additional analysis, this study explored and found the significant effects of title textual characteristics (sentiment, first-person pronouns, and vividness) on product sales, and this relationship is moderated by whether the SVA highlights product information [14]. This result indicates that creators should consider the degree of matching between the SVA content type and title expression when creating content to avoid unnecessary avoidance due to the title text of the SVA.

6.2. Theoretical Contribution

Our study contributes to existing literature in the following aspects. First, previous research has mostly focused on the effects of short video characteristics on consumer engagement or purchasing behavior, with little attention paid to the congruence characteristics of the creator's SVAs. From the perspective of creators, this study proposes and verifies the influence of title congruence of SVAs as a creation strategy, revealing its impact on advertising effectiveness and mechanism.

Second, this paper deepens our understanding of congruence characteristics. Previous studies have shown that congruence can enhance content trust and product purchase preference [32,33], while this study found that different congruence features may have varying effects on short video platforms: video-video title congruence of the same creator positively influences SVA performance, while video-product title congruence may lead to advertisement avoidance and resistance [1,51]. Dang, Saravade [95] finds that product type moderates the effect between trending content elements in SVAs and viewer engagement. Consistent with their findings, our conclusion reveals that the presentation of SVAs needs to consider content preferences, especially on short video platforms primarily aimed at social entertainment.

Third, this study reveals that the mechanism between title congruence and SVA sales performance is sociability, enriching the application of Signaling theory [96]. This finding expands the theoretical boundaries of Signaling theory in the context of short video research, indicating that sociability is not only an interaction indicator for social performance but may also become a linking mechanism driving factor for SVA sales performance.

Fourth, this study proposes creator characteristics and product characteristics as moderating variables and identifies their boundary conditions [39,69]. By introducing creation frequency and product brand as moderating variables, this study finds that the influence of title congruence on sales performance across different levels of creation frequency and product brand further emphasizes the adaptive boundaries of congruence-based creation strategies in short video platforms. This provides a foundation for future research to explore other boundary conditions, such as platform or consumer features.

Finally, this study identifies the impact of title textual features on SVA sales performance. The study finds that sentiment, first-person pronouns, and vividness in titles have a negative effect on sales performance, indicating that SVA textual features have different effects on its sales performance. This finding provides a basis for future research exploring the interaction effects of SVA text and other features.

6.3. Practical Implications

With the development of short video platforms, short video ads have increasingly become the main product-selling channel for enterprises and creators. From the perspective of creators, this study investigates and demonstrates that SVA title congruence significantly influences sales performance through the mediating role of sociability, as well as the moderating effects of creation frequency and product brand based on signaling theory. Our findings reinforce the validity of our proposed research model and offer the following practical implications for SVAs platforms:

First, creators need to maintain a reasonable level of congruence in their SVA content to establish a recognizable personal style and avoid content fatigue. This study found that creators who maintain a certain level of congruence in video-video titles can help them establish a stable creation image on the platforms and gain more recognition. However, they should also avoid content fatigue caused by excessive repetition, especially at high creation frequency, which may lead to content avoidance. Therefore, creators can tailor their congruence creation strategies, such as reusing keywords and maintaining a consistent

expression style, while innovating content themes and presentation formats to enhance consumer stickiness. Recent research has shown that content congruence significantly enhances viewer engagement in SVAs, reinforcing the importance of strategic [1].

Second, SVAs should avoid direct sale promotion, ensuring that product information is conveyed while enhancing the entertainment value and indirect nature of the SVA content. The findings of this study reveal that when video titles are highly consistent with product titles, the advertising intent is more likely to be recognized, which in turn triggers resistance, particularly on short video platforms that focus on social entertainment. Therefore, creators can adopt indirect product placement methods in the SVA creation, presenting and promoting products through storytelling, emotional appeal, or contextualization to enhance the acceptability and appeal of SVAs.

Third, introducing brand information can be an effective strategy to enhance content trust. For branded products that have already established consumer recognition and emotional attachment, their presence in SVAs can weaken advertising resistance caused by product-oriented content. Therefore, creators can promote branded products when selecting suitable products for advertising and moderately highlight brand elements in SVA contents to enhance content trust in product quality [97].

Fourth, creators should also tailor their title creation strategies to different SVA types to avoid mismatches between type and content. This study finds that sentiment, first-person pronouns, and vividness of SVA titles may weaken sales performance. Hence, when creating SVAs that do not include or indirectly mention the product, creators can try using titles with suspense or exploratory; in SVAs that include products and are aimed at product promotion, titles may focus more on a clear expression style.

Finally, although this study focuses on the Douyin platform, many insights are also transferable to its international version, TikTok, which shares comparable short video formats, integrated e-commerce functions, and algorithmic recommendations [23]. Given these structural and operational similarities, the observed relationships and mechanisms between the SVA title congruence and SVA effectiveness are expected to be suitable across TikTok and other platforms.

6.4. Limitations and Future Work

Although this study is based on real data collected from the Douyin platform, it still has the following research limitations:

First, the sample in this study focuses on creators with follower numbers larger than 100,000; this group constitutes the main content and commercial revenue on Douyin. Therefore, while the findings may not be fully generalized to micro-creators, they still provide insights into the mechanisms that drive commercial effectiveness on SVAs platforms. Future research could extend this research model by collecting data from creators with smaller follower bases to examine whether the observed relationships across different creator segments and to compare potential variations with different audience scales.

Second, the study focuses on the title of SVAs and lacks a comprehensive exploration of video content features. This study primarily focuses on the effect of title congruence (V-V and V-P title congruence) on SVA sales performance, without further analyzing how video content (such as visual style, characteristics, etc.) influences SVA performance. Future research could integrate multi-modal content analysis techniques, including video, image, and audio, to examine the overall congruence of SVA contents from a holistic perspective, thereby delving deeper into its influence and mechanism in stimulating and enhancing product sales in the context of SVAs.

Third, the selection of moderator variables is limited and fails to cover multiple heterogeneous characteristics that influence product sales. This study only selected creation

frequency and product brand as moderating variables, but product sales may also be influenced by creator features (i.e., gender, verification, and followers); product category (i.e., search-based and experiential); and individual consumer features (i.e., advertising sensitivity and purchase motivation). Future research could expand the scope of the moderating variables to reveal the conditions under which SVAs are effective.

Finally, the research data of this study is sourced from the Douyin platform in China. Considering that short video platforms in different countries or regions may vary in terms of consumer usage habits, content preferences, and cultural values, the applicability of the research conclusions across different cultural and national contexts requires further examination. Cultural and regulatory distinctions may lead to variations in viewer engagement and response to title congruence signals. Future research could conduct comparative research on other mainstream platforms, such as TikTok, YouTube, and Instagram, to examine whether the observed relationships exist across diverse cultural and platform environments. Such investigations could enhance the external validity and generalizability of the findings and provide a deeper understanding of how platform and cultural attributes shape SVA effectiveness.

Author Contributions: Conceptualization, D.H., W.J., Z.W. and R.G.; methodology, D.H. and W.J.; software, W.J. and Z.W.; formal analysis, W.J. and Z.W.; data curation, W.J. and Z.W.; writing—original draft preparation, D.H., W.J. and Z.W.; writing—review and editing, D.H., W.J., Z.W. and R.G.; funding acquisition, D.H. and W.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds for the Central Universities [grant number CXJJ-2024-464].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Datasets are available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Robustness Results for Heckman Two-Stage Test

Table A1. Results of Selection model.

Variables	Selection Model
	Propensity to Post Short Video Advertise
Creator fans	−0.729 ***
Creator expertise	0.258 ***
Video viewership	−0.194 ***
Video likes	−0.175 ***
Video comments	−0.200 ***
Video shares	0.157 ***
Video length	0.125 ***
Video age	0.068 ***
Video day of week	−0.015
Video time of day	0.001
Video title length	0.097 ***
Video tag number	0.414 ***
Month fixed effect	Yes
Creator fixed effect	Yes
Sector fixed effect	Yes

Standard errors in parentheses: *** $p < 0.01$.

Table A2. Results of observation model.

	Model 1 Sale	Model 2 Sociability	Model 3 Sale	Model 4 Sale	Model 5 Sale	Model 6 Sociability	Model 7 Sale	Model 8 Sale
V-V title congruence	0.325 ***	0.245 ***	0.234 ***	−0.091				
	(0.056)	(0.014)	(0.056)	(0.071)				
V-P title congruence					−0.166 **	−0.059 ***	−0.144 **	−0.835 ***
					(0.069)	(0.016)	(0.069)	(0.093)
Sociability			0.372 ***				0.377 ***	
			(0.018)				(0.018)	
Creation frequency				−0.306 ***				0.265 ***
				(0.058)				(0.068)
Creation frequency × V-V title congruence				0.252 ***				
				(0.076)				
Creation frequency × V-P title congruence								−0.624 ***
								(0.090)
Product brand				1.327 ***				1.018 ***
				(0.055)				(0.083)
Product brand × V-V title congruence				0.707 ***				
				(0.083)				
Product brand × V-P title congruence								1.039 ***
								(0.115)
Video day of week	0.037 *	−0.014 **	0.042 *	0.031	0.037 *	−0.013**	0.042 *	0.035
	(0.023)	(0.005)	(0.022)	(0.022)	(0.023)	(0.005)	(0.022)	(0.022)
Video time of day	−0.018	0.079 ***	−0.047 **	−0.024	−0.020	0.078 ***	−0.049 **	−0.034
	(0.024)	(0.006)	(0.024)	(0.023)	(0.024)	(0.006)	(0.024)	(0.023)
Video length	0.659 ***	0.614 ***	0.431 ***	0.483 ***	0.662 ***	0.618 ***	0.428 ***	0.492 ***
	(0.064)	(0.017)	(0.065)	(0.061)	(0.064)	(0.017)	(0.065)	(0.061)
Video age	0.239 ***	0.135 ***	0.189 ***	0.245 ***	0.240 ***	0.136 ***	0.189 ***	0.246 ***
	(0.017)	(0.004)	(0.017)	(0.016)	(0.017)	(0.004)	(0.017)	(0.016)
Video title length	0.130 ***	0.205 ***	0.054 **	0.124 ***	0.160 ***	0.224 ***	0.075 ***	0.171 ***
	(0.027)	(0.008)	(0.027)	(0.027)	(0.027)	(0.008)	(0.027)	(0.026)
Video tag number	0.237 ***	0.837 ***	−0.074 *	0.224 ***	0.248 ***	0.844 ***	−0.070 *	0.236 ***
	(0.038)	(0.010)	(0.040)	(0.036)	(0.038)	(0.011)	(0.040)	(0.036)
Product price	−0.979 ***	−0.025 ***	−0.970 ***	−0.993 ***	−0.980 ***	−0.025 ***	−0.971 ***	−0.990 ***
	(0.020)	(0.004)	(0.020)	(0.019)	(0.020)	(0.004)	(0.020)	(0.019)
Creator fans	−0.039	−1.464 ***	0.506 ***	−0.535 ***	−0.026	−1.474 ***	0.530 ***	−0.390 **
	(0.195)	(0.050)	(0.196)	(0.190)	(0.196)	(0.051)	(0.196)	(0.190)
Creator expertise	0.001	0.528 ***	−0.195 ***	0.030	−0.008	0.522 ***	−0.205 ***	0.025
	(0.027)	(0.008)	(0.028)	(0.027)	(0.027)	(0.008)	(0.028)	(0.027)

Table A2. *Cont.*

	Model 1 Sale	Model 2 Sociability	Model 3 Sale	Model 4 Sale	Model 5 Sale	Model 6 Sociability	Model 7 Sale	Model 8 Sale
Creator sex	2.901 ***	9.008 ***	−0.449	4.966 ***	2.657 ***	8.858 ***	−0.684	4.619 ***
	(0.444)	(0.106)	(0.467)	(0.448)	(0.445)	(0.106)	(0.466)	(0.443)
IMR	0.977 ***	4.303 ***	−0.623 ***	0.951 ***	0.985 ***	4.310 ***	−0.641 ***	0.971 ***
	(0.059)	(0.023)	(0.094)	(0.057)	(0.059)	(0.023)	(0.093)	(0.057)
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Creator fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	8.635 ***	12.267 ***	4.073	12.609 ***	8.685 ***	12.508 ***	3.967	10.946 ***
	(2.515)	(0.638)	(2.507)	(2.458)	(2.522)	(0.641)	(2.515)	(2.457)
N	57,192	57,192	57,192	57,192	57,192	57,192	57,192	57,192
R^2	0.551	0.905	0.555	0.581	0.551	0.905	0.555	0.581
adj. R^2	0.549	0.905	0.552	0.578	0.549	0.904	0.552	0.579

Standard errors in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B. Robustness Results for Alternative Measurement of V-V Title Congruence

Table A3. Results of alternative measurement of V-V congruence.

	Model 1 Sale	Model 2 Sociability	Model 3 Sale	Model 4 Sale	Model 5 Sale	Model 6 Sociability	Model 7 Sale	Model 8 Sale	Model 9 Sale	Model 10 Sociability	Model 11 Sale	Model 12 Sale
V-V title congruence (7)	0.208 ***	0.329 ***	0.115 **	−0.155 **								
	(0.049)	(0.020)	(0.049)	(0.063)								
V-V title congruence (21)					0.449 ***	0.432 ***	0.329 ***	−0.075				
					(0.061)	(0.024)	(0.061)	(0.075)				
V-V title congruence (28)									0.536 ***	0.490 ***	0.399 ***	0.006
									(0.065)	(0.026)	(0.065)	(0.079)
Sociability			0.283 ***				0.280 ***				0.279 ***	
			(0.011)				(0.011)				(0.011)	
Creation frequency				−0.277 ***				−0.318 ***				−0.390 ***
				(0.044)				(0.046)				(0.046)
Creation frequency × V-V title congruence (7)				0.160 ***								
				(0.054)								

Table A3. *Cont.*

	Model 1 Sale	Model 2 Sociability	Model 3 Sale	Model 4 Sale	Model 5 Sale	Model 6 Sociability	Model 7 Sale	Model 8 Sale	Model 9 Sale	Model 10 Sociability	Model 11 Sale	Model 12 Sale
Creation frequency × V-V title congruence (21)								0.238 ***				
								(0.061)				
Creation frequency × V-V title congruence (28)												0.361 ***
												(0.063)
Product brand				1.362 ***				1.269 ***				1.266 ***
				(0.055)				(0.056)				(0.056)
Product brand × V-V title congruence (7)				0.596 ***								
				(0.077)								
Product brand × V-V title congruence (21)								0.835 ***				
								(0.087)				
Product brand × V-V title congruence (28)												0.860 ***
												(0.090)
Video day of week	0.035	−0.022 **	0.042 *	0.031	0.034	−0.023 **	0.041 *	0.031	0.036	−0.022 **	0.042 *	0.032
	(0.023)	(0.009)	(0.022)	(0.022)	(0.023)	(0.009)	(0.022)	(0.022)	(0.023)	(0.009)	(0.022)	(0.022)
Video time of day	−0.030	0.028 ***	−0.038	−0.034	−0.029	0.028 ***	−0.037	−0.034	−0.030	0.027 ***	−0.038	−0.034
	(0.024)	(0.010)	(0.024)	(0.023)	(0.024)	(0.010)	(0.024)	(0.023)	(0.024)	(0.010)	(0.024)	(0.023)
Video length	0.707 ***	0.801 ***	0.476 ***	0.521 ***	0.704 ***	0.804 ***	0.475 ***	0.523 ***	0.702 ***	0.803 ***	0.475 ***	0.522 ***
	(0.064)	(0.030)	(0.064)	(0.062)	(0.064)	(0.030)	(0.064)	(0.061)	(0.064)	(0.030)	(0.064)	(0.061)
Video age	0.227 ***	0.084 ***	0.203 ***	0.232 ***	0.230 ***	0.088 ***	0.205 ***	0.235 ***	0.228 ***	0.086 ***	0.204 ***	0.234 ***
	(0.017)	(0.006)	(0.017)	(0.016)	(0.017)	(0.006)	(0.017)	(0.016)	(0.017)	(0.006)	(0.017)	(0.016)
Video title length	0.100 ***	0.031 ***	0.091 ***	0.095 ***	0.084 ***	0.023 **	0.077 ***	0.084 ***	0.077 ***	0.017	0.071 ***	0.074 ***
	(0.027)	(0.012)	(0.027)	(0.026)	(0.027)	(0.012)	(0.027)	(0.027)	(0.027)	(0.012)	(0.027)	(0.027)
Video tag number	0.063 *	0.065 ***	0.045	0.056	0.060 *	0.063 ***	0.043	0.051	0.059	0.062 ***	0.042	0.052
	(0.036)	(0.015)	(0.036)	(0.035)	(0.036)	(0.015)	(0.036)	(0.035)	(0.036)	(0.015)	(0.036)	(0.035)
Product price	−0.986 ***		−0.970 ***	−0.999 ***	−0.985 ***		−0.970 ***	−1.000 ***	−0.986 ***		−0.971 ***	−1.000 ***
	(0.020)		(0.020)	(0.019)	(0.020)		(0.020)	(0.019)	(0.020)		(0.020)	(0.019)
Creator fans	0.274	−0.085	0.293	−0.197	0.275	−0.101	0.297	−0.207	0.280	−0.098	0.301	−0.214
	(0.195)	(0.090)	(0.194)	(0.189)	(0.195)	(0.090)	(0.194)	(0.189)	(0.195)	(0.090)	(0.194)	(0.189)
Creator expertise	−0.175 ***	−0.200 ***	−0.118 ***	−0.125 ***	−0.169 ***	−0.199 ***	−0.113 ***	−0.121 ***	−0.169 ***	−0.199 ***	−0.113 ***	−0.122 ***
	(0.026)	(0.011)	(0.026)	(0.026)	(0.026)	(0.011)	(0.026)	(0.026)	(0.026)	(0.011)	(0.026)	(0.026)
Creator sex	1.909 ***	5.168 ***	0.485	3.996 ***	2.050 ***	5.241 ***	0.620	4.138 ***	2.070 ***	5.249 ***	0.644	4.091 ***
	(0.442)	(0.171)	(0.443)	(0.444)	(0.442)	(0.171)	(0.444)	(0.446)	(0.443)	(0.171)	(0.444)	(0.445)
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Creator fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	6.789 ***	3.431 ***	5.807 **	10.259 ***	6.666 ***	3.582 ***	5.652 **	10.299 ***	6.602 ***	3.550 ***	5.601 **	10.398 ***
	(2.522)	(1.163)	(2.502)	(2.454)	(2.521)	(1.159)	(2.502)	(2.453)	(2.519)	(1.159)	(2.501)	(2.453)
N	57,192	57,192	57,192	57,192	57,192	57,192	57,192	57,192	57,192	57,192	57,192	57,192
R^2	0.549	0.745	0.555	0.578	0.549	0.745	0.555	0.579	0.549	0.746	0.555	0.579
adj. R^2	0.546	0.743	0.552	0.576	0.547	0.744	0.552	0.576	0.547	0.744	0.552	0.576

Standard errors in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix C. Robustness Results for Alternative Measurement of Sociability

Table A4. Results of alternative measurement of mediator variable.

	Model 1 Sale	Model 2 Sociability Ratio	Model 3 Sale	Model 4 Sale	Model 5 Sociability Ratio	Model 6 Sale
V-V title congruence	0.356 ***	0.001 ***	0.346 ***			
	(0.056)	(0.000)	(0.057)			
V-V title congruence				−0.201 ***	−0.000 ***	−0.199 ***
				(0.069)	(0.000)	(0.069)
Sociability ratio			16.831 ***			17.480 ***
			(4.520)			(4.637)
Video day of week	0.035	−0.000 *	0.036	0.035	−0.000	0.037
	(0.023)	(0.000)	(0.023)	(0.023)	(0.000)	(0.023)
Video time of day	−0.030	−0.000 ***	−0.028	−0.031	−0.000 **	−0.030
	(0.024)	(0.000)	(0.024)	(0.024)	(0.000)	(0.024)
Video length	0.706 ***	0.001 ***	0.682 ***	0.709 ***	0.003 ***	0.683 ***
	(0.064)	(0.000)	(0.064)	(0.064)	(0.000)	(0.064)
Video age	0.228 ***	0.000 ***	0.227 ***	0.229 ***	0.000 ***	0.227 ***
	(0.017)	(0.000)	(0.017)	(0.017)	(0.000)	(0.017)
Video title length	0.090 ***	0.000	0.090 ***	0.123 ***	0.000	0.122 ***
	(0.027)	(0.000)	(0.027)	(0.027)	(0.000)	(0.027)
Video tag number	0.061 *	−0.000	0.062 *	0.073 **	0.000	0.073 **
	(0.036)	(0.000)	(0.036)	(0.036)	(0.000)	(0.036)
Product price	−0.986 ***		−0.984 ***	−0.988 ***		−0.986 ***
	(0.020)		(0.020)	(0.020)		(0.020)
Creator fans	0.282	−0.004 ***	0.342 *	0.304	−0.000 ***	0.367 *
	(0.195)	(0.001)	(0.197)	(0.196)	(0.000)	(0.198)
Creator expertise	−0.171 ***	−0.000 **	−0.169 ***	−0.183 ***	−0.000 ***	−0.180 ***
	(0.026)	(0.000)	(0.026)	(0.026)	(0.000)	(0.026)
Creator sex	1.995 ***	0.006 ***	1.893 ***	1.710 ***	0.000	1.610 ***
	(0.442)	(0.001)	(0.443)	(0.442)	(0.000)	(0.444)
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Creator fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Constant	6.635 ***	0.046 ***	5.857 **	6.612 ***	0.004 ***	5.795 **
	(2.522)	(0.007)	(2.536)	(2.530)	(0.000)	(2.544)
N	57,192	57,192	57,192	57,192	57,192	57,192
R^2	0.549	0.309	0.549	0.549	0.078	0.549
adj. R^2	0.546	0.305	0.547	0.546	0.077	0.546

Standard errors in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix D. Measurements of Sentiment, First-Person Pronoun, and Vividness

To evaluate title sentiment, we employed the "HowNet" Dictionary, a widely recognized lexical resource in Chinese natural language processing [98]. The sentiment analysis was conducted by calculating the differential frequency of positive and negative emotional words, normalized by title length. Specifically, the sentiment score was computed as follows:

$$Sentiment = \left(N_{PlusFeeling} - N_{MinusFeeling}\right) / N_{title} \tag{A1}$$

where $N_{PlusFeeling}$ represents the count of positive emotional words, $N_{MinusFeeling}$ represents the count of negative emotional words, and N_{title} denotes the words number of the title.

The first-person indicates how much the creator is inclined to adopt the first-person perspective in the short video ad. To measure the first-person, we calculated the ratio of

first-person pronouns (e.g., "I", "we", "my", "our", "me", and "us") to the total word count of the title:

$$First - person = N_{firstPerson}/N_{title} \tag{A2}$$

where $N_{firstPerson}$ represents the first-person words number, and N_{title} denotes the words number of the title.

Vividness refers to the linguistic features of a title that conveys information using rhetorical techniques. These rhetorical techniques include personal pronouns, adverbs of degree, intonation, and special sentence patterns [19]. If a title employs a rhetorical technique, it is assigned a score of 1. We then calculate the total score by summing the scores for each rhetorical technique used.

References

1. Meng, L.; Kou, S.; Duan, S.; Bie, Y. The impact of content characteristics of short-form video ads on consumer purchase behavior: Evidence from TikTok. *J. Bus. Res.* **2024**, *183*, 114874. [CrossRef]
2. Zheng, C.; Ling, S.; Cho, D.; Kim, Y. How social presence influences engagement in short video-embedded advertisements: The serial mediation effect of flow experience and advertising avoidance. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 705–724. [CrossRef]
3. Dong, X.; Liu, H.; Xi, N.; Liao, J.; Yang, Z. Short video marketing: What, when and how short-branded videos facilitate consumer engagement. *Internet Res.* **2023**, *34*, 1104–1128. [CrossRef]
4. Swani, K.; Milne, G.R.; Brown, B.P.; Assaf, A.G.; Donthu, N. What messages to post? Evaluating the popularity of social media communications in business versus consumer markets. *Ind. Mark. Manag.* **2017**, *62*, 77–87. [CrossRef]
5. Addo, P.C.; Fang, J.; Asare, A.O.; Kulbo, N.B. Customer engagement and purchase intention in live-streaming digital marketing platforms. *Serv. Ind. J.* **2021**, *41*, 767–786. [CrossRef]
6. Addo, P.C.; Akpatsa, S.K.; Nukpe, P.; Ohemeng, A.A.; Kulbo, N.B. Digital analytics approach to understanding short video advertising in digital marketing. *J. Mark. Theory Pract.* **2022**, *30*, 405–420. [CrossRef]
7. Liang, T.; Turban, E. Introduction to the special issue social commerce: A research framework for social commerce. *Int. J. Electron. Commer.* **2011**, *16*, 5–14. [CrossRef]
8. Kang, K.; Lu, J.; Guo, L.; Li, W. The dynamic effect of interactivity on customer engagement behavior through tie strength: Evidence from live streaming commerce platforms. *Int. J. Inf. Manag.* **2021**, *56*, 102251. [CrossRef]
9. Zhao, Z. Analysis on the "Douyin (TikTok) mania" phenomenon based on recommendation algorithms. *E3S Web Conf.* **2021**, *235*, 03029. [CrossRef]
10. Yang, Q.; Wang, Y.; Song, M.; Jiang, Y.; Li, Q. Sonic strategies: Unveiling the impact of sound features in short video ads on enterprise market entry performance. *J. Bus. Bus. Mark.* **2024**, *32*, 95–116. [CrossRef]
11. Lv, X.; Li, J.; Wang, Q. The dark side of recommendation algorithms in Chinese mass short video apps: Effect of perceived over-recommendation on users' cognitive dissonance and discontinuance intention. *Int. J. Hum. Comput. Int.* **2024**, *41*, 6701–6715. [CrossRef]
12. Tang, Y. Douyin platform's short-video content homogenization: An analysis of innovation dilemmas based on coordination games. *HBEM* **2025**, *60*, 108–117. [CrossRef]
13. Ge, J.; Sui, Y.; Zhou, X.; Li, G. Effect of short video ads on sales through social media: The role of advertisement content generators. *Int. J. Advert.* **2021**, *40*, 870–896. [CrossRef]
14. Cui, G.; Chung, Y.; Peng, L.; Wang, Q. Clicks for money: Predicting video views through a sentiment analysis of titles and thumbnails. *J. Bus. Res.* **2024**, *183*, 114849. [CrossRef]
15. Yoon, S.-H.; Kim, H.-W. What content and context factors lead to selection of a video clip? The heuristic route perspective. *Electron. Commer. Res.* **2019**, *19*, 603–627. [CrossRef]
16. Du, Y.; Rui, J.R.; Yu, N. How parasocial relationship and influencer-product congruence shape audience's attitude towards product placement in online videos: The mediation role of reactance. *Psychol. Res. Behav. Manag.* **2023**, *16*, 1315–1329. [CrossRef]
17. Shen, X.; Wang, J. How short video marketing influences purchase intention in social commerce: The role of users' persona perception, shared values, and individual-level factors. *Hum. Soc. Sci. Commun.* **2024**, *11*, 290. [CrossRef]
18. Haq, M.D.U.; Chiu, C.-M. Boosting online user engagement with short video endorsement content on TikTok via the image transfer mechanism. *Electron. Commer. Res. Appl.* **2024**, *64*, 101379. [CrossRef]
19. Hussain, K.; Nusair, K.; Junaid, M.; Aman, W. A two-actor model for understanding user engagement with content creators: Applying social capital theory. *Comput. Hum. Behav.* **2024**, *156*, 108237. [CrossRef]

20. Liao, J.; Chen, J. The authenticity advantage: How influencer authenticity management strategies shape digital engagement with sponsored videos. *J. Bus. Res.* **2024**, *185*, 114937. [CrossRef]
21. Xu, Y.; Chen, W.; Ow, T.T. The effects of social media posts' characteristics on customer engagement: Evidence from Wechat. *Inf. Manag.* **2023**, *60*, 103854. [CrossRef]
22. Ellis-Chadwick, F.; Chaffey, D. *Digital Marketing: Strategy, Implementation and Practice*; Pearson: London, UK, 2012.
23. Kaye, D.B.V.; Chen, X.; Zeng, J. The co-evolution of two Chinese mobile short video apps: Parallel platformization of Douyin and TikTok. *Mob. Media Commun.* **2020**, *9*, 229–253. [CrossRef]
24. Xiao, L.; Li, X.; Zhang, Y. Exploring the factors influencing consumer engagement behavior regarding short-form video advertising: A big data perspective. *J. Retail. Consum. Serv.* **2023**, *70*, 103170. [CrossRef]
25. Koh, B.; Cui, F. An exploration of the relation between the visual attributes of thumbnails and the view-through of videos: The case of branded video content. *Decis. Support Syst.* **2022**, *160*, 113820. [CrossRef]
26. Chen, H. Title Matters: Impacts of Titles on User Engagement in Short Video Platforms. Ph.D. Thesis, Singapore Management University, Singapore, 2023.
27. Hilvert-Bruce, Z.; Neill, J.; Sjöblom, M.; Hamari, J. Social motivations of live-streaming viewer engagement on twitch. *Comput. Hum. Behav.* **2018**, *84*, 58–67. [CrossRef]
28. Olson, E.L.; Thjømøe, H.M. Explaining and articulating the fit construct in sponsorship. *J. Advert.* **2011**, *40*, 57–70. [CrossRef]
29. Osgood, C.E.; Tannenbaum, P.H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.* **1955**, *62*, 42–55. [CrossRef]
30. Kim, D.Y.; Kim, H.-Y. Influencer advertising on social media: The multiple inference model on influencer-product congruence and sponsorship disclosure. *J. Bus. Res.* **2021**, *130*, 405–415. [CrossRef]
31. Eslami, S.P.; Ghasemaghaei, M. Effects of online review positiveness and review score inconsistency on sales: A comparison by product involvement. *J. Retail. Consum. Serv.* **2018**, *45*, 74–80. [CrossRef]
32. Rungruangjit, W. What drives Taobao live streaming commerce? The role of parasocial relationships, congruence and source credibility in Chinese consumers' purchase intentions. *Heliyon* **2022**, *8*, e09676. [CrossRef]
33. Park, H.; Lin, L. The effects of match-ups on the consumer attitudes toward internet celebrities and their live streaming contents in the context of product endorsement. *J. Retail. Consum. Serv.* **2020**, *52*, 101934. [CrossRef]
34. Maier, E. The negative effect of product image inconsistency on product overviews during the online product search. *Int. J. Electron. Commer.* **2019**, *23*, 110–143. [CrossRef]
35. Belanche, D.; Casaló, L.V.; Flavián, M.; Ibáñez-Sánchez, S. Understanding influencer marketing: The role of congruence between influencers, products and consumers. *J. Bus. Res.* **2021**, *132*, 186–195. [CrossRef]
36. Spence, M. Job market signaling. *Q. J. Econ.* **1973**, *87*, 355–374. [CrossRef]
37. Choi, H.S.; Ko, M.S.; Medlin, D.; Chen, C. The effect of intrinsic and extrinsic quality cues of digital video games on sales: An empirical investigation. *Decis. Support Syst.* **2018**, *106*, 86–96. [CrossRef]
38. Zhang, M.; Zhang, Y.; Zhao, L.; Li, X. What drives online course sales? Signaling effects of user-generated information in the paid knowledge market. *J. Bus. Res.* **2020**, *118*, 389–397. [CrossRef]
39. Fu, S.; Wu, Y.; Du, Q.; Li, C.; Fan, W. The secret of voice: How acoustic characteristics affect video creators' performance on bilibili. *Decis. Support Syst.* **2024**, *179*, 114167. [CrossRef]
40. Becker, M.; Gijsenberg, M.J. Consistency and commonality in advertising content: Helping or hurting? *Int. J. Res. Mark.* **2023**, *40*, 128–145. [CrossRef]
41. Hussain, M.; Islam, T.; Rehman, S.U. What you see is what you get: Assessing in-game advertising effectiveness. *J. Res. Interact. Mark.* **2022**, *17*, 527–543. [CrossRef]
42. Brown, S.; Kozinets, R.V.; Sherry, J.F. Teaching old brands new tricks: Retro branding and the revival of brand meaning. *J. Mark.* **2003**, *67*, 19–33. [CrossRef]
43. Danaher, P.J.; Bonfrer, A.; Dhar, S. The effect of competitive advertising interference on sales for packaged goods. *J. Mark. Res.* **2008**, *45*, 211–225. [CrossRef]
44. Germelmann, C.C.; Herrmann, J.-L.; Kacha, M.; Darke, P.R. Congruence and incongruence in thematic advertisement–medium combinations: Role of awareness, fluency, and persuasion knowledge. *J. Advert.* **2020**, *49*, 141–164. [CrossRef]
45. Fransen, M.L.; Fennis, B.M.; Pruyn, A.T.H. Matching communication modalities: The effects of modality congruence and processing style on brand evaluation and brand choice. *Commun. Res.* **2010**, *37*, 576–598. [CrossRef]
46. Duke, D.; Fiacconi, C.M.; Kohler, S. Parallel effects of processing fluency and positive affect on familiarity-based recognition decisions for faces. *Front. Psychol.* **2014**, *5*, 328. [CrossRef] [PubMed]
47. Jacoby, L.L. Remembering the data: Analyzing interactive processes in reading. *J. Verbal Learn. Verbal Behav.* **1983**, *22*, 485–508. [CrossRef]
48. Petty, R.E.; Cacioppo, J.T. The Elaboration Likelihood Model of Persuasion. In *Advances in Experimental Social Psychology*; Academic Press: Cambridge, MA, USA, 1986; Volume 19, pp. 123–205.

49. Lee, A.Y.; Labroo, A.A. The effect of conceptual and perceptual fluency on brand evaluation. *J. Mark. Res.* **2004**, *41*, 151–165. [CrossRef]
50. Zhang, Q.; Wang, Y.; Ariffin, S.K. Keep scrolling: An investigation of short video users' continuous watching behavior. *Inf. Manag.* **2024**, *61*, 104014. [CrossRef]
51. Elpers, J.W.; Wedel, M.; Pieters, R. Why do consumers stop viewing television commercials? Two experiments on the influence of moment-to-moment entertainment and information value. *J. Mark. Res.* **2003**, *40*, 437–453. [CrossRef]
52. Yang, B.; Zhang, X.; Cheng, X.; Xue, T. Understanding the influence mechanism of advertising avoidance from an s-o-r perspective: An empirical study based on "Qiafan" videos of Bilibili. *Electron. Commer. Res. Appl.* **2024**, *64*, 101377. [CrossRef]
53. Simola, J.; Kuisma, J.; Kaakinen, J.K. Attention, memory and preference for direct and indirect print advertisements. *J. Bus. Res.* **2020**, *111*, 249–261. [CrossRef]
54. Radach, R.; Lemmer, S.; Vorstius, C.; Heller, D.; Radach, K. Chapter 27-Eye movements in the processing of print advertisements. In *The Mind's Eye*; Hyönä, J., Radach, R., Deubel, H., Eds.; North-Holland: Amsterdam, The Netherlands, 2003; pp. 609–632.
55. Fu, W.W.; Sim, C.C. Aggregate bandwagon effect on online videos' viewership: Value uncertainty, popularity cues, and heuristics. *J. Am. Soc. Inf. Sci. Technol.* **2011**, *62*, 2382–2395. [CrossRef]
56. Leibenstein, H. Bandwagon, snob, and Veblen effects in the theory of consumers' demand. *Q. J. Econ.* **1950**, *64*, 183–207. [CrossRef]
57. Chen, L.; Yan, Y.; Smith, A.N. What drives digital engagement with sponsored videos? An investigation of video influencers' authenticity management strategies. *J. Acad. Mark. Sci.* **2022**, *51*, 198–221. [CrossRef]
58. Kronrod, A.; Huber, J. Ad wearout wearout: How time can reverse the negative effect of frequent advertising repetition on brand preference. *Int. J. Res. Mark.* **2019**, *36*, 306–324. [CrossRef]
59. Küper, A.; Krämer, N.C. Influencing factors for building social capital on live streaming websites. *Entertain. Comput.* **2021**, *39*, 100444. [CrossRef]
60. Zajonc, R.B. Mere exposure: A gateway to the subliminal. *Curr. Dir. Psychol.* **2016**, *10*, 224–228. [CrossRef]
61. Kim, K.; Lustria, M.L.; Burke, D.; Kwon, N. Predictors of cancer information overload: Findings from a national survey. *Inf. Res.* **2007**, *12*, 326.
62. Swar, B.; Hameed, T.; Reychav, I. Information overload, psychological ill-being, and behavioral intention to continue online healthcare information search. *Comput. Hum. Behav.* **2017**, *70*, 416–425. [CrossRef]
63. Cambier, F.; Poncin, I. Inferring brand integrity from marketing communications: The effects of brand transparency signals in a consumer empowerment context. *J. Bus. Res.* **2020**, *109*, 260–270. [CrossRef]
64. Upadhyay, Y.; Tripathi, A. Brand wagon effect: How brand equity eclipses the effect of eWoM on purchase intentions–Mediating role of review helpfulness. *J. Bus. Res.* **2023**, *168*, 114239. [CrossRef]
65. Ho-Dac, N.N.; Carson, S.J.; Moore, W.L. The effects of positive and negative online customer reviews: Do brand strength and category maturity matter? *J. Mark.* **2013**, *77*, 37–53. [CrossRef]
66. Ozdemir, S.; Zhang, S.; Gupta, S.; Bebek, G. The effects of trust and peer influence on corporate brand—Consumer relationships and consumer loyalty. *J. Bus. Res.* **2020**, *117*, 791–805. [CrossRef]
67. Akerlof, G.A. The market for "lemons": Quality uncertainty and the market mechanism. In *Uncertainty in Economics*; Academic Press: Cambridge, MA, USA, 1978; pp. 235–251.
68. Douyin. Douyin Data Report of 2024. Available online: https://www.douyin.com/search/2024%E6%8A%96%E9%9F%B3%E6%9C%80%E6%96%B0%E6%9C%88%E6%B4%BB%E5%AE%98%E6%96%B9%E6%95%B0%E6%8D%AE?modal_id=7423270716881128767&type=general (accessed on 25 June 2025).
69. Yuan, L.; Xia, H.; Ye, Q. The effect of advertising strategies on a short video platform: Evidence from TikTok. *Ind. Manage. Data Syst.* **2022**, *122*, 1956–1974. [CrossRef]
70. Bian, R.; Zhu, M. The relationship between ritual, personal involvement and travel intention: A study of check-in-travel on Douyin. *Am. J. Ind. Bus. Manag.* **2020**, *10*, 451–467. [CrossRef]
71. Chen, J.; Zhang, Y.; Cai, H.; Liu, L.; Liao, M.; Fang, J. A comprehensive overview of micro-influencer marketing: Decoding the current landscape, impacts, and trends. *Behav. Sci.* **2024**, *14*, 243. [CrossRef]
72. Rowley, J.; Sbaffi, L. Academics' attitudes towards peer review in scholarly journals and the effect of role and discipline. *J. Inf. Sci.* **2017**, *44*, 644–657. [CrossRef]
73. Chanmama. Douyin E-Commerce Annual Report on 2023. Available online: https://chanmama.com/school/reportDown/127.html (accessed on 22 April 2024).
74. Zhou, M.; Huang, J.; Wu, K.; Huang, X.; Kong, N.; Campy, K.S. Characterizing Chinese consumers' intention to use live e-commerce shopping. *Technol. Soc.* **2021**, *67*, 101767. [CrossRef]
75. Zhu, P.; Liu, Z.; Li, X.; Jiang, X.; Zhu, M.X. The influences of livestreaming on online purchase intention: Examining platform characteristics and consumer psychology. *Ind. Manage. Data Syst.* **2023**, *123*, 862–885. [CrossRef]
76. Boom, C.D.; Canneyt, S.V.; Demeester, T.; Dhoedt, B. Representation learning for very short texts using weighted word embedding aggregation. *Pattern Recognit. Lett.* **2016**, *80*, 150–156. [CrossRef]

77. Zhukova, A.; von Sperl, L.; Matt, C.E.; Gipp, B. Generative user-experience research for developing domain-specific natural language processing applications. *Knowl. Inf. Syst.* **2024**, *66*, 7859–7889. [CrossRef]
78. Zhou, Y.; Yang, S.; Li, Y.; Chen, Y.; Yao, J.; Qazi, A. Does the review deserve more helpfulness when its title resembles the content? Locating helpful reviews by text mining. *Inf. Process. Manag.* **2020**, *57*, 102179. [CrossRef]
79. Kim, C.; Yang, S.-U. Like, comment, and share on Facebook: How each behavior differs from the other. *Public Relat. Rev.* **2017**, *43*, 441–449. [CrossRef]
80. Yang, Y.; Zhao, Y.; Yeoh, W.; Qi, C.; Jiang, H. Exploring the interaction between streaming modes and product types in e-commerce sales. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 53. [CrossRef]
81. Baron, R.M.; Kenny, D.A. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* **1986**, *51*, 1173–1182. [CrossRef] [PubMed]
82. Jin, W.; Chen, Y.; Yang, S.; Zhou, S.; Jiang, H.; Wei, J. Personalized managerial response and negative inconsistent review helpfulness: The mediating effect of perceived response helpfulness. *J. Retail. Consum. Serv.* **2023**, *74*, 103398. [CrossRef]
83. Li, R.; Lu, Y.; Ma, J.; Wang, W. Examining gifting behavior on live streaming platforms: An identity-based motivation model. *Inf. Manag.* **2021**, *58*, 103406. [CrossRef]
84. Syrdal, H.A.; Myers, S.; Sen, S.; Woodroof, P.J.; McDowell, W.C. Influencer marketing and the growth of affiliates: The effects of language features on engagement behavior. *J. Bus. Res.* **2023**, *163*, 113875. [CrossRef]
85. Unsloth, A.I.; Han, D.; Han, M. *Unsloth*. Github. 2025. Available online: http://github.com/unslothai/unsloth (accessed on 25 June 2025).
86. Chang, S.-T.; Wu, J.-J. A content-based metric for social media influencer marketing. *Ind. Manag. Data Syst.* **2023**, *124*, 344–360. [CrossRef]
87. Chew, R.; Bollenbacher, J.; Wenger, M.; Speer, J.; Kim, A. LLM-assisted content analysis: Using large language models to support deductive coding. *arXiv* **2023**. [CrossRef]
88. Li, Y.; Kim, H.-j.; Do, B.; Choi, J. The effect of emotion in thumbnails and titles of video clips on pre-roll advertising effectiveness. *J. Bus. Res.* **2022**, *151*, 232–243. [CrossRef]
89. Packard, G.; Moore, S.G.; McFerran, B. (I'm) happy to help (you): The impact of personal pronoun use in customer-firm interactions. *J. Mark. Res.* **2018**, *55*, 541–555. [CrossRef]
90. Chen, Y.; Zhou, S.; Jin, W.; Chen, S. Investigating the determinants of medical crowdfunding performance: A signaling theory perspective. *Internet Res.* **2022**, *33*, 1134–1156. [CrossRef]
91. Stern, B.B. Who talks advertising? Literary theory and narrative "point of view". *J. Advert.* **1991**, *20*, 9–22. [CrossRef]
92. Zhang, Z.; Zhang, L. Most significant impact on consumer engagement: An analytical framework for the multimodal content of short video advertisements. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 54. [CrossRef]
93. Albert, N.; Ambroise, L.; Valette-Florence, P. Consumer, brand, celebrity: Which congruency produces effective celebrity endorsements? *J. Bus. Res.* **2017**, *81*, 96–106. [CrossRef]
94. Luo, C.; Hasan, N.A.M.; Zamri Bin Ahmad, A.M.; Lei, G. Influence of short video content on consumers purchase intentions on social media platforms with trust as a mediator. *Sci. Rep.* **2025**, *15*, 16605. [CrossRef]
95. Dang, A.; Saravade, S.; Spradlin, J.; Nworah, P.; Linstead, A.; Munger, R. Should brands utilize trending content elements for their short-form advertisements? *Mark. Manag. J.* **2025**, *35*, 1–22. [CrossRef]
96. Yang, Z.; Dongqi, S. How are brands associated by users in short videos-a study on the mechanism of user associations with brand placements in short videos based on signal theory. *PLoS ONE* **2025**, *20*, e0316905. [CrossRef]
97. Choi, H.S.; Ayaburi, E.W. Effect of corporate reputation and coopetitive branding strategy on digital products sales: Evidence from digital video game industry. *Int. J. Inf. Manag.* **2026**, *86*, 102975. [CrossRef]
98. Dong, Z.; Dong, Q. HowNet-a hybrid language and knowledge resource. In Proceedings of the International Conference on Natural Language Processing and Knowledge Engineering, Beijing, China, 26–29 October 2003; pp. 820–824. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

healthcare

MDPI

Perspective

Incompleteness of Electronic Health Records: An Impending Process Problem Within Healthcare

Varadraj Gurupur [1,2,*], Sahar Hooshmand [3], Deepa Fernandes Prabhu [1,4], Elizabeth Trader [1,5] and Sanket Salvi [1,2]

1 Center for Decision Support Systems and Informatics, University of Central Florida, Orlando, FL 32816, USA;deepa.fernandesprabhu@ucf.edu (D.F.P.); elizabeth.trader@ucf.edu (E.T.); sanket.salvi@ucf.edu (S.S.)
2 School of Global Health Management and Informatics, University of Central Florida, Orlando, FL 32816, USA
3 Department of Computer Science, California State University, Dominguez Hills, Carson, CA 90747, USA; hooshmand@csudh.edu
4 School of Modeling, Simulation and Training, University of Central Florida, Orlando, FL 32816, USA
5 Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL 32816, USA
* Correspondence: varadraj.gurupur@ucf.edu

Abstract

Background: The digitization of health records was expected to improve data quality and accessibility, yet incompleteness remains a widespread challenge that undermines clinical care, interoperability, and downstream analytics. Problem: Evidence shows that missing and under-recorded elements in electronic health records (EHRs) are largely driven by process gaps across patients, providers, technology, and policy—not solely by technical limitations. Objective: This perspective integrates conceptual foundations of incompleteness, synthesizes cross-country evidence, and examines process-level drivers and consequences, with an emphasis on how missingness propagates bias in AI and machine learning systems. Contribution: We present a unifying taxonomy, highlight complementary approaches (e.g., Record Strength Score, distributional testing, and workflow studies), and we propose a pragmatic agenda for mitigation through technical, organizational, governance, and patient-centered levers. Conclusions: While EHR incompleteness cannot be fully eliminated, it can be systematically mitigated through standards, workflow redesign, patient engagement, and governance—essential steps toward building safe, equitable, and effective learning health systems.

Keywords: electronic health records; incompleteness; data quality; missing data; interoperability; bias; clinical documentation; machine learning; process improvement; global health

1. Introduction

Electronic health records (EHRs) have become a cornerstone of modern healthcare, providing critical infrastructure for clinical decision-making, research, and health system management. Global adoption has grown rapidly: many high-income countries now have national or regional EHR systems in place (often exceeding 50–70% adoption), whereas adoption in low- and lower-middle-income countries remains substantially lower, often in the range of 10–40%, depending on region [1]. In the United States, nearly 96% of hospitals and 78% of office-based physicians now use certified EHR systems [2]. Similarly, the European Union has committed to creating a European Health Data Space by 2025, aiming to improve cross-border interoperability of EHRs [3].

Despite this widespread implementation, the incompleteness of EHR data remains a critical, unresolved challenge. Empirical studies reveal that many EHR datasets have

substantial missingness: in some settings, 30–40% of variables may be missing more than half their expected values, especially for laboratory or social determinants data, and overall missingness rates for key clinical variables often fall in the 15–30% range, depending on context [4–6]. Missing or fragmented EHR records have been linked to an increased risk of medication errors, redundant testing, misdiagnoses, and worse health outcomes, especially among marginalized populations. Evidence shows that discontinuities in patient records degrade predictive model fairness for racial and ethnic minorities [7], while missing data and documentation gaps contribute to diagnostic inaccuracies and bias in clinical decision-making [8,9]. As health systems increasingly rely on EHR data for predictive analytics, population health management, and machine learning applications, the risks associated with incompleteness are magnified.

While prior research has described aspects of data incompleteness, most existing work remains fragmented. Earlier contributions have focused either on conceptual frameworks [10], technical completeness metrics [11], or specific methodological biases [12]. However, a unifying perspective that situates incompleteness as a systemic process problem spanning patients, providers, technology, and policy has been largely absent. Furthermore, the global dimension of this issue manifested in differences across health systems, governance structures, and digital divides has not been adequately integrated into prior reviews. The novelty of this article lies in the following contributions:

- It provides an integrative framing of EHR incompleteness across six interdependent domains: definitions and dimensions, types of missingness, research contributions, process gaps, implications, and mitigation strategies with future directions.
- It synthesizes empirical findings from multiple countries, moving beyond single-region perspectives.
- It highlights both the statistical foundations and the sociotechnical drivers of incompleteness.
- It emphasizes the ethical and policy consequences of incomplete records, linking data quality to equity and governance.

In parallel with these academic and operational developments, major strategic and regulatory bodies have also advanced initiatives to improve the completeness, interoperability, and governance of electronic health records. For instance, the European Commission's European Health Data Space initiative seeks to harmonize standards and enhance the quality and accessibility of health data across member states, promoting cross-border interoperability and secondary data use for research and innovation [3,13,14]. Similarly, the U.S. Food and Drug Administration (FDA) has emphasized real-world data reliability and electronic health record usability as part of its broader digital health modernization efforts, reinforcing the importance of structured and complete data capture for regulatory decision-making. These activities reflect growing international momentum toward policy frameworks that prioritize data completeness as a foundation for safe, equitable, and interoperable digital health systems [13,15].

1.1. Perspective Design and Scope

As this work is presented as a perspective, rather than a systematic review, its objective is to synthesize key conceptual, empirical, and policy-oriented developments related to EHR incompleteness. The selection of literature emphasized peer-reviewed studies published between 2013 and 2025 that addressed data completeness, interoperability, bias, and governance within electronic health record systems. Foundational theoretical frameworks were included to establish conceptual grounding, complemented by recent computational and policy-driven studies that reflect current trends. The perspective was designed to integrate diverse viewpoints—spanning technical, organizational, patient-centered, and policy domains—into a coherent argument highlighting incompleteness

as a systemic process issue. This narrative and integrative design was chosen to encourage cross-disciplinary dialogue and to identify pragmatic directions for future research and governance.

The remainder of this article is structured as follows: Section 2 develops the conceptual foundations of incompleteness, including definitions, mathematical representations, and a taxonomy of missingness. Section 3 reviews key research contributions, bridging conceptual, statistical, and computational perspectives. Section 4 synthesizes cross-country evidence, underscoring the global scope of the problem. Section 5 identifies process gaps across patients, providers, technology, and policy. Section 6 analyzes the implications of incompleteness for clinical care, operations, research, and ethics. Section 7 outlines mitigation strategies and future directions, spanning technical, organizational, policy, and patient-centered solutions. Finally, Section 8 presents conclusions and authorial reflections.

1.2. Definitions and Dimensions

The IEEE Standards Association defines incompleteness in electronic health records (EHRs) as the absence of a required data element that should exist for a patient encounter [16]. Building upon this, Weiskopf et al. [10] proposed four operational dimensions:

- Documentation—whether a data element is recorded at all.
- Breadth—the extent to which all relevant data types are captured.
- Density—the frequency of data entries over time.
- Predictive value—the ability of a record to support reliable inference about outcomes.

As shown in Table 1, each dimension of completeness can be formalized mathematically. Documentation and breadth quantify whether essential data types are present, while density emphasizes longitudinal data capture across encounters. Predictive value links record completeness to downstream inferential strength, bridging data quality and clinical utility. Together, these metrics offer a structured foundation for evaluating EHR completeness beyond qualitative assessment [10,11].

Table 1. Mathematical representations of EHR completeness dimensions [10,11].

Dimension	Formal Expression
Documentation	$D_i = \begin{cases} 1 & \text{if element } i \text{ is documented} \\ 0 & \text{otherwise} \end{cases}$, where i indexes required elements.
Breadth	$B = \frac{\sum_{i=1}^{k} D_i}{k}$, where k is the total number of required data types.
Density	$Den = \frac{\sum_{j=1}^{t} E_j}{t}$, where E_j is the number of entries at observation point j, and t is the total number of periods.
Predictive Value	$PV = f(\hat{Y}(X), Y)$, where $\hat{Y}(X)$ is the predicted outcome based on record X, and Y is the true outcome (measured by AUROC, accuracy, or similar).

To concretize the four completeness dimensions, we computed RSS on a de-identified MIMIC-IV subset of type 2 diabetes encounters (n = 1000). As shown in Table 2, two medical informaticians selected "required" items for diabetes management; inter-rater agreement was high ($\kappa = 0.91$). Component values were normalized to $[0, 1]$ and aggregated with equal weights:

$$\text{RSS} = 0.25 \cdot \text{Doc} + 0.25 \cdot \text{Bre} + 0.25 \cdot \text{Den} + 0.25 \cdot \text{PV}.$$

To enable cross-domain comparability, components were z-normalized within-dataset prior to aggregation.

Table 2. Required items and decision rules for the diabetes example (consensus panel of two informaticians).

Required Item	Decision Rule (per Encounter)	Clinical Rationale
HbA1c	≥1 structured HbA1c result recorded	Glycemic control indicator for T2D management.
Systolic/Diastolic BP	≥1 structured SBP/DBP pair recorded	Hypertension risk assessment and therapy titration.
ICD-10 E11.xx Comorbidity	Any E11.xx code present	Confirms T2D diagnosis context and supports classification consistency.
Medication Adherence Note	Any adherence entry (structured or coded NLP)	Indicates therapy effectiveness and disease progression monitoring.

Table 3 provides an illustrative calculation for three encounters, showing the individual component scores (documentation, breadth, density, and predictive value) and their aggregated RSS values. The visual distribution of these normalized components across encounters is presented in Figure 1, where the bar profiles highlight variation in data completeness among patient records. To ensure cross-domain comparability, each component was z-normalized within its dataset prior to aggregation, and the "required" item set was defined through expert consensus.

Table 3. Example component scores for three encounters (normalized to $[0, 1]$). Den = entries per 90 days; PV = standardized feature importance in 6-month readmission model.

Encounter	Doc	Bre	Den	PV	RSS Calculation
P001	0.75	0.75	0.60	0.70	$0.25(0.75 + 0.75 + 0.60 + 0.70) = 0.70$
P002	1.00	1.00	0.45	0.62	$0.25(1.00 + 1.00 + 0.45 + 0.62) = 0.77$
P003	0.50	0.50	0.30	0.55	$0.25(0.50 + 0.50 + 0.30 + 0.55) = 0.46$
Mean ± SD	0.75 ± 0.25	0.75 ± 0.25	0.45 ± 0.15	0.62 ± 0.08	0.64 ± 0.16

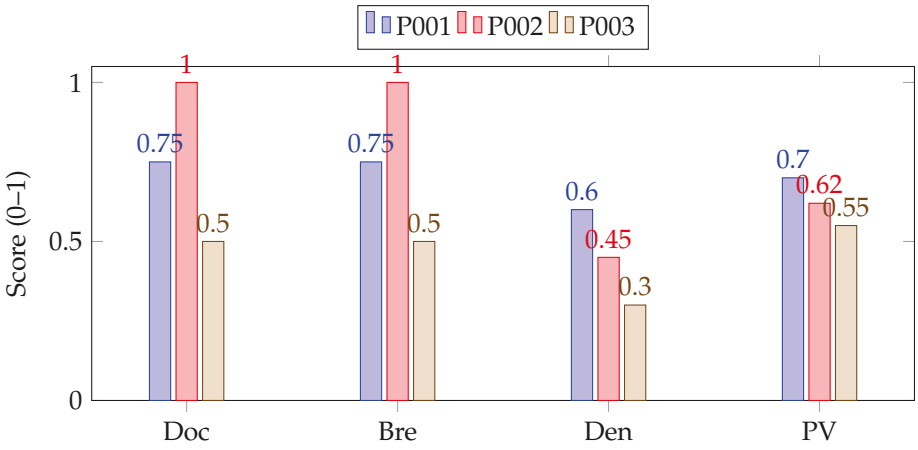


Figure 1. Component profiles for three example encounters (normalized). Equal-weight aggregation yields RSS $= 0.25(\text{Doc} + \text{Bre} + \text{Den} + \text{PV})$.

1.3. Types of Missingness

Rubin's taxonomy [17] distinguishes between three canonical forms of missingness:

- Missing Completely at Random (MCAR):

$$P(M \mid X, Y) = P(M) \tag{1}$$

Missingness is independent of both observed and unobserved data.

- Missing at Random (MAR):

$$P(M \mid X, Y) = P(M \mid X) \tag{2}$$

Missingness depends only on observed data X.

- Missing Not at Random (MNAR):

$$P(M \mid X, Y) \neq P(M \mid X) \tag{3}$$

Missingness depends on the unobserved data itself.

Here, M is the missingness indicator ($M = 1$ if missing), X denotes observed data, and Y unobserved values. Identifying the missingness mechanism is crucial for selecting appropriate imputation or modeling strategies in EHR analyses.

As this work offers a conceptual perspective, validation was discussed, rather than empirically performed. We illustrate how simulated data and the Kolmogorov–Smirnov test can support distinguishing random from systematic missingness [18]. Expert interpretation, aligned with prior frameworks [10,12], further informs practical classification. A conceptual illustration summarizing the complementary roles of statistical and expert reasoning in identifying MCAR, MAR, and MNAR patterns is presented in Table 4. This reflective approach emphasizes combining statistical reasoning with expert judgment to guide the robust identification of missingness mechanisms across EHR contexts.

Table 4. Conceptual illustration of missingness mechanism identification using statistical and expert approaches.

Mechanism	Statistical Cue (K-S Test Outcome)	Interpretive Insight (Expert Reasoning)
MCAR	Distributions of observed vs. missing values show no significant difference ($p > 0.05$)	Likely random omission unrelated to clinical or system factors.
MAR	Deviations linked to observed covariates (e.g., age, visit type)	Missingness associated with measurable variables in workflow or population.
MNAR	K-S test reveals significant skew; pattern persists after conditioning ($p < 0.05$)	Data absence related to unrecorded or sensitive factors (e.g., diagnosis disclosure).

1.4. A Taxonomy for EHR Incompleteness

Incompleteness extends beyond data properties to include causal pathways and downstream effects. Figure 2 presents an integrated taxonomy comprising six domains: Data Dimensions (documentation, breadth, density, predictive value); Types of Missingness (MCAR, MAR, MNAR); Causes (human factors, administrative constraints, technological limitations, patient-generated data challenges); Detection and Quantification Approaches (RSS, Kolmogorov–Smirnov test [18], graph theory [19], pattern recognition [20]); Process Gaps (patient, provider, technology, and administrative levels); and Implications (clinical decision risk, interoperability challenges such as HL7/FHIR, regulatory issues, and bias in machine learning models).

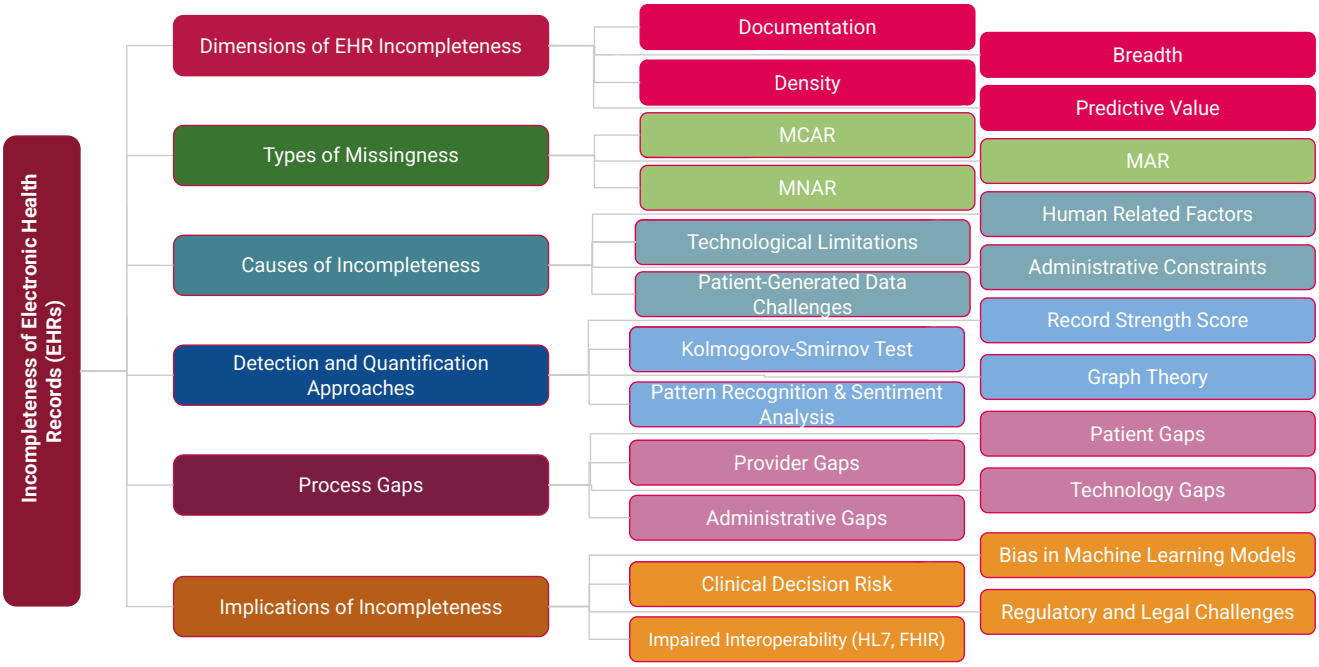


Figure 2. Integrated taxonomy of factors contributing to incompleteness in EHRs, spanning data dimensions, missingness types, causes, detection and quantification methods, process gaps, and implications.

This taxonomy underscores that incompleteness is not merely a technical flaw but a systemic process problem with profound clinical, organizational, and ethical consequences.

2. Key Research Contributions: Complementary Lenses

Research into EHR incompleteness has progressed through several distinct phases. Early work focused on establishing conceptual frameworks, defining the dimensions of completeness, and highlighting the structural gaps in health records. This conceptual grounding provided the language for later operationalization. Building on these foundations, subsequent studies introduced quantifiable measures that allowed completeness to be evaluated more systematically. Among these, the Record Strength Score (RSS) proposed by Nasir et al. [11] offered a simple yet powerful way to quantify data completeness as the ratio of recorded to required elements within a patient's clinical context:

$$RSS = \frac{N_{\text{present}}}{N_{\text{required}}} \tag{4}$$

where N_{present} is the number of required elements documented, and N_{required} is the total expected number of elements. This measure was later extended by Nasir et al. [21] to examine disparities across patient subgroups, demonstrating that incompleteness often reflects underlying inequities in healthcare access and utilization.

As the field matured, more rigorous statistical methods were employed to characterize incompleteness patterns. For instance, Gurupur et al. [22] applied the Kolmogorov–Smirnov (K–S) test [18] to detect whether missingness was random or systematic. The K–S statistic is defined as follows:

$$D = \sup_x |F_1(x) - F_2(x)| \tag{5}$$

where D is the maximum distance between the observed empirical cumulative distribution, $F_1(x)$, and the expected distribution, $F_2(x)$. By identifying deviations from expected distributions, this method provides insight into whether incompleteness arises from systematic process gaps, rather than random variation.

More recently, advanced computational approaches have reframed incompleteness as a learnable pattern. Ontology-driven validation has been used to detect structural errors in clinical model design [23], and ML classifiers have been built to predict which physicians will omit documentation in hospital EHR systems [24]. Real-world studies also continue to uncover missingness patterns across demographic, behavioral, and health history fields [25]. These developments emphasize that incompleteness is not merely a technical issue but also a behavioral and systemic one, requiring methods that integrate statistical rigor with contextual understanding. As shown in Table 5, key research contributions on EHR incompleteness are summarized chronologically.

Table 5. Representative research contributions on EHR incompleteness, arranged chronologically.

Study/Line of Work	Core Contribution	Limitations/Notes
Weiskopf et al. (2013) [10]	Defined four dimensions of completeness (documentation, breadth, density, predictive value); provided a conceptual framework for measuring completeness.	Primarily conceptual; lacks quantitative tools for operational assessment across large datasets.
Nasir et al. (2016) [11]	Introduced the Record Strength Score (RSS) to mathematically quantify completeness as the ratio of present to required elements.	Context-specific; requires predefined lists of "required" data elements.
Nasir et al. (2019) [21]	Extended RSS to explore disparities in record completeness across demographics; identified higher missingness in vulnerable populations.	Focused on utilization data; does not generalize across all EHR types.
Wilcox et al. (2021) [19]	Classified missingness into MCAR, MAR, MNAR; applied graph theory to model data flows and identify gaps in smart healthcare environments.	Graph-theoretic models require strong assumptions; practical deployment may be limited in heterogeneous EHR systems.
Gurupur et al. (2021) [20]	Applied distributional approaches (Kolmogorov–Smirnov test) and machine learning to predict incompleteness patterns.	Site-specific validation; computational complexity increases with scale.
Gurupur et al. (2022) [26]	Modeled incompleteness as a learnable pattern using ontologies and recurrent neural networks; linked provider sentiment to completeness.	Requires specialized datasets; generalizability remains underexplored.
Tsiampalis & Panagiotakos (2023) [12]	Reviewed methodological issues in EHR-based epidemiological studies, focusing on biases from missingness.	Narrative scope; lacks empirical validation.
Gurupur et al. (2024) [16]	Provided a standards-based narrative review; emphasized the importance of universal definitions and quality benchmarks.	Descriptive; does not provide quantitative validation.
Fraser et al. (2024) [27]	Compared data quality (completeness, accuracy, timeliness) between paper and EHR implementations.	Context-specific to a limited health system; generalization needed.
Kim et al. (2024) [8]	Highlighted challenges and opportunities in EHR-based research, emphasizing data quality and missingness.	Broad review; does not propose formal metrics.
Mugauri et al. (2025) [28]	Examined a decade of EHR implementations in Sub-Saharan Africa, identifying structural causes of incompleteness.	Regional focus; limited generalizability.

As summarized in Table 6, research on EHR incompleteness has progressed along a chronological trajectory from early conceptual frameworks to advanced computational modeling. Initial work by Weiskopf et al. (2013) [10] established key dimensions of completeness, followed by the introduction of quantitative measures such as the Record Strength Score (RSS) by Nasir and colleagues (2016, 2019) [11,21]. More recent studies have applied graph theory, statistical distribution tests, and machine learning to characterize and predict incompleteness [19,20,22]. Reviews [12,27] and empirical studies [8,28] from 2023 onward reflect both global perspectives and emerging standards, underscoring that incompleteness is a multidimensional issue requiring conceptual, methodological, and policy-driven solutions.

Recent global studies reinforce that incompleteness is not a U.S.-centric phenomenon [12]. Reviews from Europe [12], empirical evaluations in Africa [28], and comparative analyses of EHR versus paper-based records [27] demonstrate that data gaps persist even in advanced health systems and are compounded in resource-limited settings. Taken together, these contributions reveal a multidimensional research trajectory that naturally motivates a broader synthesis of cross-country evidence, which we present in the next section.

While previous studies have examined EHR incompleteness from isolated viewpoints—such as conceptual frameworks [10], completeness metrics like the Record Strength Score (RSS) [11], or methodological bias analyses [12]—our framework provides a unifying perspective that integrates these approaches within a process-oriented model. Specifically, it bridges technical and behavioral dimensions by situating data incompleteness within patient, provider, technology, and policy workflows. This comparison underscores that, unlike earlier studies that primarily focused on measurement or statistical aspects alone, our framework advances the field by linking incompleteness to systemic process gaps and offering actionable pathways for mitigation through multi-level coordination.

Table 6. Representative contributions on EHR incompleteness (approach, contribution, and typical limitations).

Study/Line of Work	Core Contribution	Typical Limitations/Notes
Conceptual frameworks	Definitions; dimensions of completeness; quality measures	May lack operational metrics; variable generalizability
Graph/flow modeling & missingness classes	Characterizes data sources and flows; clarifies MAR/MCAR/MNAR	Requires assumptions; limited by observability of processes
Record Strength Score (RSS)	Quantifies completeness relative to a required set	Dependent on context-specific "required" element lists
Distributional testing/ML	Detects systematic patterns; links to sentiments/workflows	Needs robust validation; sensitive to site-specific idiosyncrasies
Standards/benchmarks	Shared definitions, reporting, and evaluation guidance	Adoption/implementation gaps across vendors/sites

3. Global Evidence of EHR Incompleteness

While much of the early research on EHR incompleteness originated in the United States, recent work shows that the problem is global in scope. Studies across diverse regions demonstrate that missingness arises not only from technical limitations but also from structural, organizational, and socio-political factors. Table 7 summarizes representative evidence across countries and regions, illustrating that incompleteness is a universal challenge with context-specific manifestations.

Table 7. Cross-country evidence of EHR incompleteness.

Region/Country	Context/Dataset	Key Incompleteness Patterns	Reference(s)
United States	Medicare, Medicaid, multi-institutional EHRs	Under-recording of chronic conditions, incomplete problem lists, disparities across demographics.	[10,21]
United States	Multi-institutional predictive modeling	Data discontinuity amplifies racial-ethnic disparities; incomplete longitudinal capture reduces fairness.	[7]
United Kingdom	NHS general practice and hospital records	Missing ethnicity, lifestyle factors (e.g., smoking), and mental health severity scores.	[12]
Nordics (Sweden, Denmark)	National health registries	Underreporting of comorbidities, registry update delays, incomplete linkage across systems.	[12]
Germany	Surgical departments post-EMR adoption	Some documentation fields improved, but perioperative and lab data remained incomplete.	[29]
Belgium	Seven-hospital case study	Basic measures (height, weight) inconsistently recorded across sites and specialties.	[30]
China	Urban vs. rural hospitals	Documentation disparities; fragmented records across health tiers.	[31]
India	Mixed public–private settings	Low EHR adoption; structural incompleteness due to partial digitization.	[32]
South Korea/Japan	Advanced hospital EHRs	Missing behavioral and lifestyle data; limited PGHD integration.	[33]
Kuwait	Public hospitals	Quantitative evaluation of EHR safety revealed substantial missing safety-related fields.	[34]
Sub-Saharan Africa	HIV/maternal health programs	Missing labs, imaging, and continuity data due to infrastructure constraints.	[28]
Latin America (Brazil, Mexico)	National/state health systems	Inconsistent reporting; missing vaccination and chronic disease data.	[35]

As Table 8 demonstrates, incompleteness in EHRs is a universal challenge, but it manifests differently, depending on health system maturity and resources. In high-income countries such as the United States and United Kingdom, incompleteness often arises from workflow limitations, underreporting in structured fields, or a lack of interoperability across vendor systems. In contrast, low- and middle-income countries face structural incompleteness due to limited EHR adoption, partial digitization, and infrastructural challenges such as unreliable connectivity. Hybrid contexts like Latin America highlight the role of policy and governance, with regional variation in reporting standards contributing to data gaps.

Concrete case studies reinforce these findings. In Germany, Wurster et al. (2024) [29] showed that, even after EMR adoption, perioperative and laboratory documentation remained incomplete. In Belgium, Declerck et al. (2024) [30] demonstrated that routine measurements such as height and weight were often missing or inconsistently recorded across hospitals. In the United States, Huang et al. (2023) [4] revealed that incomplete longitudinal data continuity can amplify racial–ethnic disparities in predictive modeling. In Kuwait, AlHussainan et al. (2025) [34] identified critical gaps in safety-related fields across public hospitals. These examples highlight that incompleteness persists even in resource-rich settings and has direct consequences for clinical care, equity, and research validity.

To enable comparability across health systems with differing EHR maturity and data definitions, prior studies highlight the utility of schema alignment through HL7/FHIR mappings and the Observational Medical Outcomes Partnership (OMOP) Common Data Model [15]. Reported incompleteness rates in the reviewed literature were interpreted as

relative indicators after accounting for contextual factors such as data model scope and reporting granularity.

Recognizing its global scope is essential for developing mitigation strategies that are context-sensitive while also guided by universal standards. The next section explores the specific process gaps at the patient, provider, technology, and administrative levels, that drive incompleteness within healthcare workflows.

Table 8. Cross-country evidence of EHR incompleteness (illustrative).

Country/Region	Context/Dataset	Salient Incompleteness Pattern	Reported/Expected Impact
United States	Claims + EHR hybrids	Under-recorded chronic condition severity; problem list gaps	Misclassification; reimbursement issues; model bias
United Kingdom	GP/EHR registries	Missing ethnicity, smoking, mental health fields	Limits epidemiology; equity analyses
Nordics (SE/DK)	National registries	Underreporting of comorbidities; lag in updates	Risk adjustment drift; outcome misestimation
China	Mixed hospital systems	Rural/urban documentation disparities	Fragmented longitudinal care; biased analytics
India	Low EHR penetration	Structural incompleteness; paper-to-digital transitions	Limited secondary use; safety risks
Sub-Saharan Africa	Pilots/verticals	Missing labs/imaging; connectivity/storage constraints	Clinical delays; surveillance blind spots
Latin America (BR/MX)	State/regional systems	Inconsistent reporting across jurisdictions	Program evaluation bias; data linkage failures

4. Process Gaps Driving Incompleteness

As presented in Figure 3 and Table 9, incompleteness in electronic health records emerges from interacting process gaps across four domains: patients, providers, technology, and administrative/policy environments. Recent studies demonstrate how each domain contributes distinctive patterns of missingness and fragmentation.

Table 9. Summary of key process gaps contributing to EHR incompleteness.

Domain	Representative Findings (Cross-Country Evidence)	Key Contributing Factors and Gaps
Patient Level	Missing data correlate with demographic, socioeconomic, and access disparities (U.S., UK, Africa).	Non-disclosure of sensitive information; variability in patient-generated health data (PGHD); digital divide; language and cultural barriers.
Provider Level	Incomplete documentation and workflow burden prevalent in Europe and the U.S.	Clinician fatigue; high documentation load; workflow misalignment; insufficient validation and feedback mechanisms.
Technology Level	Interoperability and usability gaps in Belgium, Germany, and Asia impact completeness.	Poor interface usability (UI/UX); inconsistent data standards (HL7/FHIR); fragmented storage; limited PGHD integration.
Policy/Administrative Level	Governance and regulatory variation shape documentation quality (Kuwait, Canada, Denmark).	Lack of standardized protocols; weak accountability; misaligned incentives; limited enforcement of documentation quality.

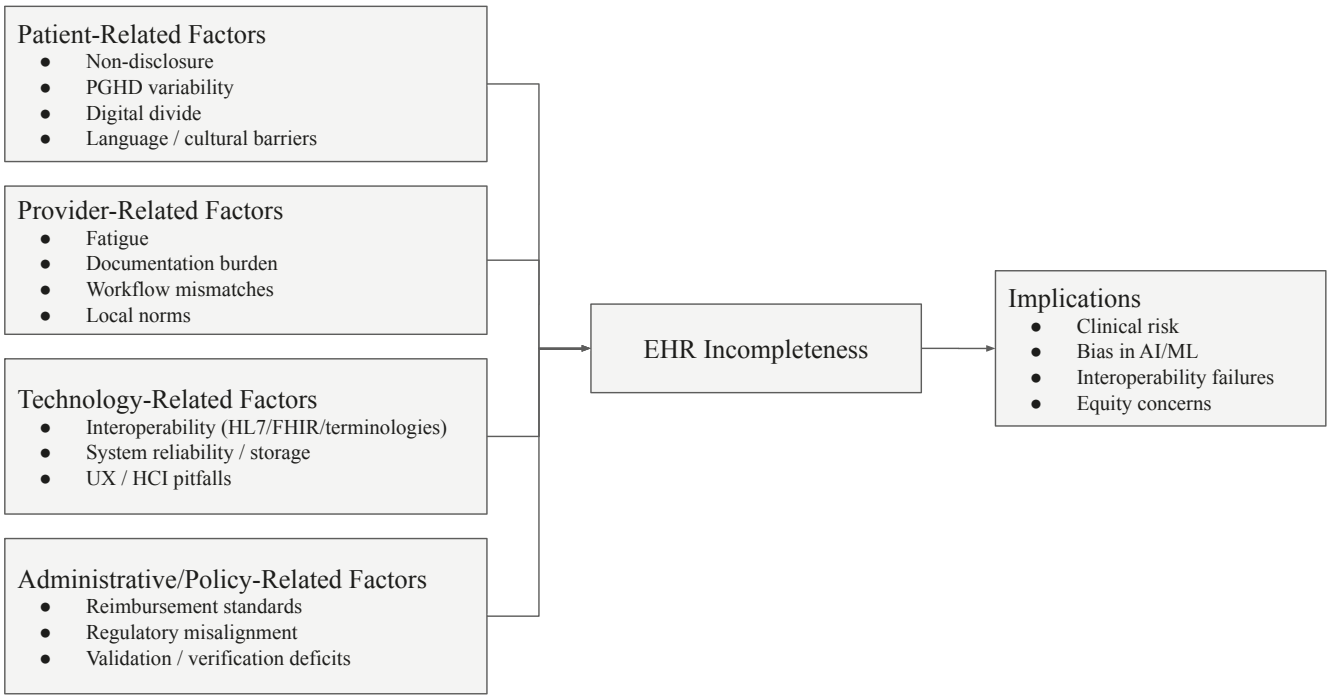


Figure 3. Conceptual model of interacting process gaps (patient, provider, technology, and administrative/policy) leading to EHR incompleteness.

4.1. Patient-Related Factors

Patient-level dynamics shape the continuity, access, and completeness of records. For instance, Teeple et al. (2023) found that missing data patterns vary by race and that these disparities degrade clinical prediction fairness and accuracy [36]. Choy et al. (2024) further found that patients face significant barriers in digital health access—including poor connectivity and limited digital literacy—translating into missing engagement data [37]. Conderino et al. (2024) showed that patients with cost-constrained or delayed care experienced systematically worse data reliability, as self-reported conditions were often absent or underrepresented in the record [38].

Earlier interviews also revealed human-related contributors such as the non-disclosure of sensitive conditions, fatigue, and breakdowns in human-to-human communication (e.g., language or cultural barriers), particularly in emergency care [39]. Together, these findings underscore that patient-driven gaps reflect both systemic inequities and day-to-day human behavior.

4.2. Provider-Related Factors

Provider workflows, burden, and usability remain central to documentation completeness. Wurster et al. (2024) reported that perioperative and laboratory fields remained incomplete in German surgical departments, even after EMR adoption [29]. Sloss et al. (2024) showed that the redesign of templates, workflow optimization, and the use of medical scribes can reduce burdens and improve documentation [40]. Olakotan et al. (2025) found that poor interface usability and confusing navigation led providers to skip or leave fields incomplete [41].

Qualitative evidence echoes these findings: interviews highlighted human error (deliberate and inadvertent), fatigue, and a lack of verification/validation of inputs as persistent bottlenecks [42]. Medication errors arising from missing or incomplete drug histories exemplify how such gaps can have catastrophic consequences [43].

4.3. Technology-Related Factors

System design, interoperability, and UX strongly influence which data are captured and which are omitted. Declerck et al. (2024) found that basic variables like height and weight were inconsistently recorded across Belgian hospitals due to system defaults and inconsistent requirements [30]. Olakotan et al. (2025) noted that suboptimal UIs (lag, non-intuitive layouts) drive incomplete entries [41]. Das et al. (2024) proposed the CSSDH ontology to integrate social determinants with continuity of care standards, showing that standardized ontologies can reduce incompleteness [44].

Interviews further revealed interoperability challenges as a major barrier: fragmentation between EHR systems leads to missing critical data during care transitions, limiting the effectiveness of HL7/FHIR frameworks [19,22]. Issues with unreliable storage (e.g., lack of RAID-based redundancy) can also generate incomplete or inaccessible patient data [45,46]. Patient-generated health data, when unstructured or lacking standards, compounds the problem.

4.4. Administrative/Policy-Related Factors

Policy and governance shape the incentives and constraints for completeness. Al-Hussainan et al. (2025) identified safety-critical omissions in Kuwait's public hospitals linked to weak verification and oversight [34]. Kim et al. (2025) provide a framework for digital health equity that mandates embedding equity across all phases of digital health governance—from planning and procurement to implementation and monitoring—to prevent systemic access and documentation gaps [47]. In parallel, Hu et al. (2024) find that individuals who are uninsured, less educated, or living in rural or deprived areas are significantly more likely to experience lower EHR completeness tied to access-related policy barriers [48].

Earlier analyses revealed that a lack of standardized protocols and weak accountability perpetuate bottlenecks in data collection, leaving critical fields unvalidated or defaulted to inaccurate values [22]. These deficits in policy and verification systems enable systemic bias in records, which then propagates into analytics and decision-support models.

Together, the four domains illustrated in Figure 3 show that EHR incompleteness arises from a combination of patient access and behavior, provider burden and usability, technology design and interoperability, and policy and administrative oversight. Addressing these demands multi-level interventions: equity in access, usability-centered design, the standardization of technology, and coherent policy frameworks.

5. Implications of Incompleteness

The consequences of incomplete electronic health records (EHRs) extend beyond data management to directly influence clinical decision-making, operational efficiency, research validity, and ethical governance. Incompleteness introduces risks that permeate multiple levels of the healthcare ecosystem. This section synthesizes the key implications across four domains: clinical, operational and interoperability, research and machine learning, and ethical and policy considerations.

5.1. Clinical Implications

From a clinical perspective, incompleteness compromises the accuracy and timeliness of decision-making. Recent evidence shows that information omissions and documentation defects ("EHRrors") are common and clinically consequential, encompassing data that are incorrectly recorded or never captured when they should have been [49]. Studies also associate EHR design and use with patient-safety outcomes: incomplete or poorly surfaced histories contribute to medical errors and delay appropriate therapy, whereas the better visibility

of prior results can reduce unnecessary tests and downstream risk [50,51]. In high-acuity settings such as emergency departments, the cognitive load created by EHR workarounds and fragmented information further elevates the likelihood of missed data, misinterpretation, and treatment delays [52]. Collectively, these findings emphasize that EHR incompleteness is not merely a data-quality nuisance but a direct threat to patient safety.

5.2. Operational and Interoperability Implications

Operationally, missing or inconsistently recorded data drive inefficiencies—duplicated diagnostics, repeated imaging, and reimbursement delays—that increase costs without improving outcomes. Empirical evidence shows that, when clinicians can reliably access prior records across facilities, redundant testing decreases, highlighting how completeness supports the streamlining of care pathways [53]. At the systems level, interoperability frameworks (HL7/FHIR) presuppose complete, well-structured data to enable safe exchange; incomplete payloads degrade handoffs and continuity, especially during transitions of care [54]. Real-world implementations underscore this point: large-scale FHIR deployments reveal recurring "pain points" (e.g., missing required elements, vocabulary mismatches) that impair data flow and undermine the value of interoperability investments [55]. Finally, the documentation burden and misaligned EHR workflows remain persistent operational liabilities that correlate with incomplete entries and downstream rework [56].

5.3. Research and Machine Learning Implications

Incomplete EHR data compromise the reliability of biomedical research and the performance of AI models trained on routine care data. Bias can originate from systematic gaps in key covariates (e.g., race/ethnicity), with downstream effects on model calibration, subgroup performance, and external validity [36,57]. State-of-the-art surveys of fair machine learning in health repeatedly identify missingness and data incompleteness as primary sources of algorithmic bias and performance disparities [58,59]. Beyond fairness, incomplete data reduce statistical power, complicate causal inference, and increase reliance on unverifiable assumptions in imputation pipelines; consensus guidance stresses that mitigation must be coupled to explicit characterization of the missingness mechanism and subgroup impact [60]. Relatedly, disparities in how clinicians and staff engage with EHR systems across patient populations can themselves induce patterned missingness, further threatening research generalizability [6].

5.4. Ethical and Policy Implications

Ethically, EHR incompleteness raises questions of justice, beneficence, and accountability because the harms of missing data are unevenly distributed and can amplify existing disparities in access and outcomes. Reviews on digital health equity argue that governance should explicitly target data completeness and representativeness as prerequisites for equitable care and AI deployment [47,61]. At the policy level, experts emphasize that semantic interoperability and standard alignment are necessary but insufficient without assurances that the underlying records are complete and fit for reuse [62,63]. In AI-enabled workflows, ethical analyses call for institutional accountability to detect, report, and remedy completeness-related bias before models influence care decisions [64–66]. Altogether, ethical and policy considerations underscore that completeness is not merely a technical metric; it is foundational to fair, trustworthy, and legally defensible digital health systems.

6. Mitigation Strategies and Future Directions

With the conceptual foundations, process gaps, and broad implications of incompleteness having been outlined, it is critical to turn toward solutions. Addressing incompleteness requires a multi-level approach that integrates technical, organizational, policy, and patient-centered interventions. Each of these domains contributes distinct but complementary mechanisms to reduce missingness, harmonize workflows, and build trust in electronic health records (EHRs). This section synthesizes emerging strategies and future directions across these four domains.

6.1. Technical Solutions

Technical innovations offer the most direct pathway to mitigating incompleteness in EHRs. Real-time data quality assessment systems that flag missing, inconsistent, or implausible entries at the point of documentation can substantially reduce downstream errors. Recent studies demonstrate that automated completeness checks integrated into EHR workflows improve both data capture and clinical usability [67–69]. Beyond rule-based validation, machine learning methods have been developed to predict likely missing values and harmonize heterogeneous inputs across systems. For example, natural language processing pipelines now extract structured information from unstructured clinical notes, partially closing gaps in structured fields [70,71].

Standardization remains equally critical. Despite the wide adoption of HL7/FHIR, the lack of uniform implementation undermines the consistency of captured data. Ongoing initiatives have emphasized maturing FHIR profiles and strengthening mapping between terminologies such as SNOMED CT and LOINC to reduce structural incompleteness during interoperability [15,72]. Finally, technical advances in federated learning and privacy-preserving record linkage are showing promise in integrating fragmented patient records across institutions without compromising security. For example, Kho et al. (2020) used PPRL in the All of Us Research Program to detect care fragmentation across health provider organizations [73]. Similarly, recent frameworks such as FED-EHR (Wani et al. 2025) and reviews of methodological advances in FL (Zhang et al. 2024) demonstrate that federated models can maintain performance while preserving privacy [74,75]. Altogether, these methods help mitigate longitudinal incompleteness by allowing information linkage and learning across distributed datasets.

6.2. Organizational Solutions

Organizational strategies are essential to address the human and workflow dimensions of incompleteness. Evidence consistently shows that high documentation burden and poorly aligned workflows are leading contributors to missing or incomplete EHR data. Interventions such as team-based documentation, medical scribes, and task redistribution have been associated with improved completeness and reduced clinician fatigue [76]. Workflow redesign, usability improvements, and reduced click/navigation burden likewise appear to improve documentation completeness and reduce burnout [77,78].

Training and education represent another critical organizational strategy. Studies reveal that clinicians and ancillary staff often lack sufficient training in structured data entry, resulting in incomplete or inconsistent records [56,61]. Structured onboarding, refresher courses, and role-specific training have demonstrated measurable gains in data completeness across hospital units.

Finally, organizational culture plays an important role. Institutions that emphasize quality improvement and data stewardship, supported by leadership commitment and feedback mechanisms, show higher completeness rates compared to those without such frameworks [79]. Equally important is addressing patient-facing workflows, where initia-

tives such as improving digital literacy support, patient portals, and structured patient-generated health data entry have shown promise in reducing missingness [80,81].

Together, these findings suggest that organizational solutions are indispensable complements to technical interventions, ensuring that completeness is embedded into day-to-day clinical practice rather than treated as an afterthought.

6.3. Policy and Governance Solutions

Policy and governance interventions provide the structural backbone for improving EHR completeness across health systems. Unlike technical or organizational efforts that focus on workflows or tools, governance frameworks establish accountability, incentives, and standards that shape long-term data quality. Recent work emphasizes that national and regional governance structures significantly influence data completeness. For example, a systematic review of digital health governance models demonstrated that countries with mandated reporting standards and centralized oversight (e.g., Denmark, Estonia) achieved higher levels of record completeness compared to systems with fragmented governance [13]. Similarly, regulatory policies enforcing the mandatory coding of diagnoses and medications were associated with significant improvements in EHR completeness in France and Canada [82]. Another crucial area involves the alignment of reimbursement incentives with documentation quality. Studies indicate that, when reimbursement policies reward accurate and complete documentation, providers demonstrate higher compliance with structured EHR data entry [83].

Conversely, poorly designed incentives risk promoting "check-box" behavior without meaningful improvements in data quality. Governance frameworks must also address interoperability and equity. A multicountry 2024 study reported that the absence of coordinated regulatory mandates for HL7/FHIR adoption perpetuates fragmentation and missingness, especially in cross-border care [14,15]. At the same time, the failure to incorporate equity-driven governance exacerbates disparities, as underrepresented populations often face higher levels of incomplete data capture [84].

Together, these findings indicate that policy and governance solutions must extend beyond compliance checklists. Effective governance should embed standardization, accountability, financial alignment, and equity considerations as cornerstones of completeness. Without such frameworks, technical and organizational interventions risk remaining piecemeal, unable to address systemic incompleteness across healthcare ecosystems.

6.4. Patient-Centered Solutions

Patients play a pivotal role in mitigating EHR incompleteness, especially through engagement, transparency, and the contribution of patient-generated health data (PGHD). A growing body of evidence suggests that, when patients are empowered as active participants in their health data journey, record completeness and accuracy improve substantially.

One promising approach is the use of patient portals and mobile health applications that allow patients to review and supplement their records. A multicenter trial demonstrated that providing patients with structured opportunities to correct medication lists and allergy information via portals significantly increased data completeness while reducing discrepancies during clinical encounters [85,86]. Similarly, mobile apps integrated with EHRs have been shown to improve the reporting of chronic disease management metrics, such as blood glucose and blood pressure, filling critical gaps in longitudinal monitoring [87,88]. Another important dimension is addressing health literacy and digital divide barriers. Studies show that patients with limited digital literacy are less likely to contribute reliable PGHD, perpetuating disparities in data completeness [89,90]. Targeted

training programs and simplified user interfaces can help mitigate this issue, ensuring more equitable patient participation in data contribution.

Finally, involving patients in data governance and quality oversight has gained traction as a mechanism for accountability. A 2025 study reported that incorporating patient representatives in data quality review committees improved trust and identified gaps that clinicians and administrators often overlooked [91]. In addition, expanding the integration of PGHD through wearables and remote monitoring devices provides richer datasets, although the challenges of standardization and validation remain [92,93]. Taken together, patient-centered solutions highlight that data completeness is not solely a technical or institutional responsibility. Patients themselves, when supported with the right tools, literacy, and governance frameworks, become key contributors to more accurate and equitable EHR systems.

Table 10 consolidates the wide range of mitigation strategies discussed in this section. Technical solutions address structural and interoperability gaps, while organizational interventions reduce documentation burden and embed data quality into workflows. Policy and governance frameworks provide accountability and standardization, and patient-centered initiatives extend completeness by empowering individuals to contribute and verify their own health data. Together, these strategies emphasize that no single intervention is sufficient; sustained improvements require coordinated action across all four domains.

Table 10. Summary of mitigation strategies for addressing EHR incompleteness across technical, organizational, policy, and patient-centered domains.

Domain	Example Strategies	Key Notes/Limitations
Technical Solutions	- Real-time data quality checks. - NLP extraction from clinical notes. - Federated learning for fragmented data. - Enhanced FHIR/SNOMED/LOINC mapping.	Effective for structural gaps but require robust IT infrastructure and standardization across vendors.
Organizational Solutions	- Team-based documentation. - Medical scribes. - Structured training programs. - Workflow redesign with clinician input.	Reduce documentation burden and improve completeness, but sustainability depends on culture and leadership support.
Policy and Governance	- National reporting standards. - Regulatory incentives for complete coding. - Alignment of reimbursement with quality. - Equity-focused governance models.	Provide long-term accountability but risk "checkbox compliance" if not paired with monitoring and equity safeguards.
Patient-Centered Solutions	- Patient portals with correction rights. - Mobile health apps for PGHD. - Health literacy interventions. - Patient involvement in quality oversight.	Improve completeness and trust, but effectiveness depends on digital literacy, access, and the usability of tools.

7. Final Remarks

The synthesis presented in this article underscores that electronic health record (EHR) incompleteness is not solely a technical problem but a systemic process challenge that cuts across patient behavior, provider workflows, technology design, and policy environments. Mitigation, therefore, demands multi-level coordination—combining standardization, usability, equity-oriented governance, and patient empowerment. These reflections aim to consolidate the conceptual and practical implications discussed throughout, setting the stage for the concluding section.

Limitations

As a perspective, this work does not provide a systematic or quantitative evaluation of all available studies on EHR incompleteness. The synthesis is inherently interpretive and relies on selective integration of conceptual, empirical, and policy literature. While efforts were made to ensure broad coverage and balanced representation, some regional or domain-specific nuances may not be fully captured. Future research could build on this framework through systematic reviews or meta-analyses that quantitatively assess completeness patterns and mitigation outcomes across settings.

8. Conclusions

This perspective has highlighted that incompleteness in electronic health records (EHRs) is not merely a technical artifact but a systemic process problem, arising at the intersection of patient behaviors, provider workflows, technological limitations, and policy structures. Across sections, we demonstrated how incompleteness compromises clinical decision-making, weakens interoperability, biases research and machine learning models, and raises serious ethical and governance concerns. From an authorial standpoint, our central argument is that EHR incompleteness must be treated as a foundational barrier to the promise of digital health. It is not sufficient to pursue interoperability or advanced analytics without first ensuring that the underlying data are consistently captured, validated, and contextualized. Our synthesis suggests that mitigation requires a multi-level approach: technical advances such as automated quality checks and federated integration; organizational strategies that reduce documentation burden and embed usability; governance models that align incentives and mandate standards; and patient-centered solutions that empower individuals to contribute and verify their health data. While absolute completeness is unattainable, systematic improvement is both feasible and urgent. By embedding completeness into the design of health information systems and aligning it with equity, accountability, and patient safety, we can shift EHRs from fragmented repositories toward reliable, learning health infrastructures. As researchers and practitioners, we see this as an opportunity: addressing incompleteness is not simply a corrective task but a necessary step toward realizing the transformative potential of digital health for safe, equitable, and globally relevant healthcare systems.

Author Contributions: Conceptualization, V.G.; methodology, V.G., D.F.P., and S.H.; validation, V.G. and S.S.; formal analysis, E.T. and S.S.; investigation, S.H. and D.F.P.; resources, V.G.; data curation, E.T. and S.S.; writing—original draft preparation, S.S., V.G., and S.H.; writing—review and editing, D.F.P. and E.T.; visualization, S.S.; supervision, V.G.; project administration, V.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AUROC	Area Under the Receiver Operating Characteristic Curve
EHR	Electronic Health Record

FHIR	Fast Healthcare Interoperability Resources
HL7	Health Level Seven
ICD-10	International Classification of Diseases, 10th Revision
LOINC	Logical Observation Identifiers Names and Codes
MCAR	Missing Completely at Random
MAR	Missing at Random
MNAR	Missing Not at Random
ML	Machine Learning
NLP	Natural Language Processing
PGHD	Patient-Generated Health Data
RSS	Record Strength Score
SNOMED CT	Systematized Nomenclature of Medicine—Clinical Terms
UX/HCI	User Experience/Human–Computer Interaction

References

1. Shapiro, L.M.; Kamal, R.N. Implementation of Electronic Health Records During Global Outreach: A Necessary Next Step in Measuring and Improving Quality of Care. *J. Hand Surg.* **2022**, *47*, 279–283. [CrossRef]
2. Office of the National Coordinator for Health Information Technology. *Office-Based Physician Electronic Health Record Adoption*; Health IT Quick-Stat #50; Office of the National Coordinator for Health Information Technology: Washington, DC, USA, 2022.
3. European Commission. European Health Data Space: Regulation Proposal. COM/2022/197 Final. 2022. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022PC0197 (accessed on 2 October 2025).
4. Zhou, Y.; Shi, J.; Stein, R.; Liu, X.; Baldassano, R.N.; Forrest, C.B.; Chen, Y.; Huang, J. Missing data matter: An empirical evaluation of the impacts of missing EHR data in comparative effectiveness research. *J. Am. Med. Inform. Assoc.* **2023**, *30*, 1246–1256. [CrossRef] [PubMed]
5. Li, J.; Yan, X.S.; Chaudhary, D.; Avula, V.; Mudiganti, S.; Husby, H.; Shahjouei, S.; Afshar, A.; Stewart, W.F.; Yeasin, M.; et al. Imputation of missing values for electronic health record laboratory data. *Npj Digit. Med.* **2021**, *4*, 147. [CrossRef] [PubMed]
6. Cesare, N.; Were, L.P.O. A multi-step approach to managing missing data in time and patient variant electronic health records. *BMC Res. Notes* **2022**, *15*, 64. [CrossRef] [PubMed]
7. Kar, S.; Bessette, L.G.; Wyss, R.; Kesselheim, A.S.; Lin, K.J. The impact of electronic health record discontinuity on prediction modeling. *PLoS ONE* **2023**, *18*, e0287985. [CrossRef]
8. Kim, M.K.; Rouphael, C.; McMichael, J.; Welch, N.; Dasarathy, S. Challenges in and Opportunities for Electronic Health Record-Based Data Analysis and Interpretation. *Gut Liver* **2024**, *18*, 201–208. [CrossRef]
9. Perets, O.; Stagno, E.; Yehuda, E.B.; McNichol, M.; Celi, L.A.; Rappoport, N.; Dorotic, M. Inherent Bias in Electronic Health Records: A Scoping Review of Sources of Bias. *medRxiv* **2024**. [CrossRef]
10. Weiskopf, N.G.; Weng, C. Defining and measuring completeness of electronic health records for secondary use. *J. Biomed. Inform.* **2013**, *46*, 830–836. [CrossRef]
11. Nasir, A.; Gurupur, V.; Liu, X. A new paradigm to analyze data completeness of patient data. *Appl. Clin. Inform.* **2016**, *7*, 745–764. [CrossRef]
12. Tsiampalis, T.; Panagiotakos, D.B. Methodological issues of the electronic health records' use in the context of epidemiological investigations, in light of missing data. *BMC Med. Res. Methodol.* **2023**, *23*, 20. [CrossRef]
13. Bente, B.; Beerlage-de Jong, N.; Verdaasdonk, R.; van Gemert-Pijnen, J. Unveiling Practical Insights of eHealth Implementation in Europe: A Grey Literature Review on Legal, Ethical, Financial, and Technological (LEFT) Considerations. *Front. Digit. Health* **2025**, *7*, 1575620. [CrossRef] [PubMed]
14. Bossenko, I.; Randmaa, R.; Piho, G.; Ross, P. Interoperability of Health Data Using FHIR Mapping Language: Transforming HL7 CDA to FHIR with Reusable Visual Components. *Front. Digit. Health* **2024**, *6*, 1480600. [CrossRef] [PubMed]
15. Tabari, P.; Costagliola, G.; De Rosa, M.; Boeker, M. State-of-the-Art Fast Healthcare Interoperability Resources (FHIR)-Based Data Model and Structure Implementations: Systematic Scoping Review. *JMIR Med. Inform.* **2024**, *12*, e58445. [CrossRef] [PubMed]
16. Gurupur, V.P.; Vu, G.; Mayya, V.; King, C. The Need for Standards in Evaluating the Quality of Electronic Health Records and Dental Records: A Narrative Review. *Big Data Cogn. Comput.* **2024**, *8*, 168. [CrossRef]
17. Rubin, D.B. Inference and missing data. *Biometrika* **1976**, *63*, 581–592. [CrossRef]
18. Massey, F.J. The Kolmogorov-Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.* **1951**, *46*, 68–78. [CrossRef]
19. Wilcox, C.; Djahel, S.; Giagos, V. Identifying the main causes of medical data incompleteness in the smart healthcare era. In Proceedings of the 2021 International Symposium on Networks, Computers and Communications (ISNCC), Dubai, United Arab Emirates, 31 October–2 November 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6. [CrossRef]

20. Gurupur, V.P.; Shelleh, M. Machine Learning Analysis for Data Incompleteness (MADI): Analyzing the Data Completeness of Patient Records Using a Random Variable Approach. *IEEE Access* **2021**, *9*, 95994–96001. [CrossRef]
21. Nasir, A.; Liu, X.; Gurupur, V.; Qureshi, Z. Disparities in patient record completeness with respect to the health care utilization project. *Health Inform. J.* **2019**, *25*, 401–416. [CrossRef]
22. Gurupur, V.P.; Abedin, P.; Hooshmand, S.; Shelleh, M. Analyzing the Data Completeness of Patients' Records Using a Random Variable Approach to Predict the Incompleteness of Electronic Health Records. *Appl. Sci.* **2022**, *12*, 10746. [CrossRef]
23. Martínez-Costa, C.; Schulz, S. Validating EHR clinical models using ontology patterns. *J. Biomed. Inform.* **2017**, *76*, 124–137. [CrossRef]
24. Al Habib, A.F.; Alharthi, H.M. Prediction of Electronic Health Record Documentation Compliance Using Machine Learning. *Perspect. Health Inf. Manag.* **2023**, *20*, 1e.
25. Tiwari, T.; Hughes, J.; Granick, L.; Holmes, J.H.; Pacheco, J.A.; Jones, J.; Bush, W.S. Evaluating the completeness of electronic health records in a large national biobank. *JAMIA Open* **2025**, *8*, ooae004. [CrossRef]
26. Gurupur, V.P.; Gurupur, V.P.; Wan, T.T.H.; Rudraraju, R.R.; Kulkarni, S.A. Can the theories of information and communication channels be used to explain the complexities associated with transformation of data into information, and information to knowledge? *J. Integr. Des. Process Sci.* **2022**, *27*, 59–69. [CrossRef]
27. Fraser, H.S.F.; Mugisha, M.; Bacher, I.; Ngenzi, J.L.; Seebregts, C.; Umubyeyi, A.; Condo, J. Factors Influencing Data Quality in Electronic Health Record Systems in 50 Health Facilities in Rwanda and the Role of Clinical Alerts: Cross-Sectional Observational Study. *JMIR Public Health Surveill.* **2024**, *10*, e49127. [CrossRef] [PubMed]
28. Mugauri, H.D.; Chimsimbe, M.; Shambira, G.; Shamhu, S.; Nyamasve, J.; Munyanyi, M.; Gongora, R.; Zizhou, S.; Gabida, M.; Makurumidze, R. A decade of designing and implementing electronic health records in Sub-Saharan Africa: A scoping review. *Glob. Health Action* **2025**, *18*, 2492913. [CrossRef]
29. Wurster, F.; Herrmann, C.; Beckmann, M.; Cecon-Stabel, N.; Dittmer, K.; Hansen, T.; Jaschke, J.; Köberlein-Neu, J.; Okumu, M.-R.; Pfaff, H.; et al. Differences in changes of data completeness after the implementation of an electronic medical record in three surgical departments of a German hospital–a longitudinal comparative document analysis. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 258. [CrossRef]
30. Declerck, J.; Vandenberk, B.; Deschepper, M.; Colpaert, K.; Cool, L.; Goemaere, J.; Bové, M.; Staelens, F.; De Meester, K.; Verbeke, E.; et al. Building a Foundation for High-Quality Health Data: Multihospital Case Study in Belgium. *JMIR Med. Inform.* **2024**, *12*, e60244. [CrossRef]
31. Liang, J.; Li, Y.; Zhang, Z.; Shen, D.; Xu, J.; Zheng, X.; Wang, T.; Tang, B.; Lei, J.; Zhang, J. Adoption of Electronic Health Records (EHRs) in China During the Past 10 Years: Consecutive Survey Data Analysis and Comparison of Sino-American Challenges and Experiences. *J. Med. Internet Res.* **2021**, *23*, e24813. [CrossRef]
32. Gandhi, A.P.; Soundappan, K. Perception towards electronic health records & uptake of digital health IDs among the urban residents in northern India: A mixed methods study. *Indian J. Med. Res.* **2024**, *160*, 51–60. [CrossRef]
33. Shin, S.Y. Current status and future direction of digital health in Korea. *Korean J. Physiol. Pharmacol.* **2019**, *23*, 311–315. [CrossRef]
34. AlHussainan, A.; Alhuwail, D. Quantitative Evaluation of Electronic Health Record Safety in Public Hospitals in Kuwait. *JMIR Med. Inform.* **2025**, *13*, e70782. [CrossRef]
35. Barbalho, I.M.; Fernandes, F.; Barros, D.M.; Paiva, J.C.; Henriques, J.; Morais, A.H.; Coutinho, K.D.; Coelho Neto, G.; Chioro, A.; Valentim, R.A. Electronic Health Records in Brazil: Prospects and Technological Challenges. *Front. Public Health* **2022**, *10*, 963841. [CrossRef] [PubMed]
36. Teeple, S.; Smith, A.; Toerper, M.; Levin, S.; Halpern, S.; Badaki-Makun, O.; Hinson, J. Exploring the impact of missingness on racial disparities in machine-learning clinical prediction. *JAMIA Open* **2023**, *6*, ooad107. [CrossRef] [PubMed]
37. Choy, M.A.; O'Brien, K.; Barnes, K.; Sturgiss, E.A.; Rieger, E.; Douglas, K. Evaluating the Digital Health Experience for Patients in Primary Care: Mixed Methods Study. *J. Med. Internet Res.* **2024**, *26*, e50410. [CrossRef] [PubMed] [PubMed Central]
38. Conderino, S.E.; Haneuse, S.; Roy, J.A.; Moodie, E.E.; Buchanan, A.L. Addressing Information Biases Within Electronic Health Records: Missing Data and Causal Inference Frameworks. *JMIR Med. Inform.* **2024**, *12*, e58085. [CrossRef]
39. Khairat, S.; Coleman, C.; Ottmar, P.; Jayachander, D.I.; Bice, T.; Carson, S.S. Association of Electronic Health Record Use With Physician Fatigue and Efficiency. *JAMA Netw. Open* **2020**, *3*, e207385. Erratum in *JAMA Netw Open* **2020**, *3*, e2013153. [CrossRef]
40. Sloss, E.A.; Abdul, S.; Aboagyewah, M.A.; Beebe, A.; Kendle, K.; Marshall, K.; Rosenbloom, S.T.; Rossetti, S.; Grigg, A.; Smith, K.D.; et al. Toward Alleviating Clinician Documentation Burden: A Scoping Review of Burden Reduction Efforts. *Appl. Clin. Inform.* **2024**, *15*, 446–455. [CrossRef] [PubMed]
41. Olakotan, O.; Samuriwo, R.; Ismaila, H.; Atiku, S. Usability Challenges in Electronic Health Records: Impact on Documentation Burden and Clinical Workflow: A Scoping Review. *J. Eval. Clin. Pract.* **2025**, *31*, e70189. [CrossRef]
42. Tawfik, D.; Sebok-Syer, S.S.; Bragdon, C.; Brown-Johnson, C.; Winget, M.; Bayati, M.; Shanafelt, T.; Profit, J. Qualitative Verification of Machine Learning-Based Burnout Predictors in Primary Care Physicians: An Exploratory Study. *Appl. Clin. Inform.* **2025**, *16*, 1031–1040. [CrossRef]

43. Wittich, C.M.; Burkle, C.M.; Lanier, W.L. Medication errors: An overview for clinicians. *Mayo Clin. Proc.* **2014**, *89*, 1116–1125. [CrossRef]
44. Das, S.; Naskar, D.; González, S.R. CSSDH: An Ontology for Social Determinants of Health to Operational Continuity of Care Data Interoperability. In *Intelligent Data Engineering and Automated Learning—IDEAL 2024*; Julian, V., Camacho, D., Yin, H., Alberola, J.M., Nogueira, V.B., Novais, P., Tallón-Ballesteros, A., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2025; Volume 15346, pp. 489–500. [CrossRef]
45. Thantilage, R.D.; Le-Khac, N.A.; Kechadi, M.T. Healthcare data security and privacy in Data Warehouse architectures. *Inform. Med. Unlocked* **2023**, *39*, 101270. [CrossRef]
46. Shojaei, P.; Vlahu-Gjorgievska, E.; Chow, Y.W. Security and Privacy of Technologies in Health Information Systems: A Systematic Literature Review. *Computers* **2024**, *13*, 41. [CrossRef]
47. Kim, K.K.; Backonja, U. Digital health equity frameworks and key concepts. *J. Am. Med. Inform. Assoc.* **2025**, *32*, 932–945. [CrossRef] [PubMed]
48. Hu, Q.; Yao, Y.; Han, J.; Yang, X.T.; Parton, J. Examining the existing usage gap of electronic health records in the United States: A study of National Health Survey. *SSM—Popul. Health* **2023**, *25*, 101577. [CrossRef]
49. Kharko, A.; Hägglund, M.; Angelova, D.; Scott Duncan, T.; Hagström, J.; Hansford, K.; Hunt, J.; Trolle Lagerros, Y.; Locher, C.; McMillan, B.; et al. Prevalence and types of errors in the electronic health record: Protocol for a mixed systematic review. *BMJ Open* **2025**, *15*, e098241. [CrossRef]
50. Nijor, S.; Rallis, G.; Lad, N.; Gokcen, E. Patient Safety Issues From Information Overload in Electronic Medical Records. *J. Patient Saf.* **2022**, *18*, e999–e1003. [CrossRef]
51. Bowman, S. Impact of electronic health record systems on information integrity: Quality and safety implications. *Perspect. Health Inf. Manag.* **2013**, *10*, 1c.
52. Colicchio, T.K.; Cimino, J.J.; Del Fiol, G. Unintended Consequences of Nationwide Electronic Health Record Adoption: Challenges and Opportunities in the Post-Meaningful Use Era. *J. Med. Internet Res.* **2019**, *21*, e13313. [CrossRef]
53. Ben-Assuli, O.; Sagi, D.; Amador Nelke, S.; Leshno, M.; Ziv, A.; Ironi, A. The impact of electronic health records on the ordering of medical tests. *Isr. J. Health Policy Res.* **2025**, *14*, 18. [CrossRef]
54. Chou, E.; Sockolow, P.S. Transitions of Care: Completeness of the Interoperability Data Standard for Communication from Home Health Care to Primary Care. *AMIA Annu Symp Proc.* **2022**, *2021*, 295–304.
55. Ceusters, W.; Rabenberg, M. Semantic Difficulties in FHIR 'Conditions'. *Stud. Health Technol. Inform.* **2025**, *327*, 7–11. [CrossRef]
56. Moy, A.; Shea, C.M.; Weston, M.J.; Applegate, M.; Salazar, A.; Pines, J.M.; Ward, M.J. Understanding the perceived role of electronic health records and workflow fragmentation on clinician documentation burden in emergency departments. *J. Am. Coll. Emerg. Physicians Open* **2023**, *4*, e12956. [CrossRef]
57. Tsalidis, A.; Bharadwaj, L.; Shen, F.X. Standardization and accuracy of race and ethnicity data: Equity implications for medical AI. *PLoS Digit. Health* **2025**, *4*, e0002467. [CrossRef] [PubMed]
58. Feng, Q.; Du, M.; Zou, N.; Hu, X. Fair Machine Learning in Healthcare: A Survey. *IEEE Trans. Artif. Intell.* **2025**, *6*, 493–507. [CrossRef]
59. Ueda, D.; Kakinuma, T.; Fujita, S.; Kamagata, K.; Fushimi, Y.; Ito, R.; Matsui, Y.; Nozaki, T.; Nakaura, T.; Fujima, N.; et al. Fairness of Artificial Intelligence in Healthcare: Review and Emerging Agenda. *Artif. Intell. Med.* **2023**, *136*, 102456. [CrossRef]
60. Raman, S.R.; Hammill, B.G.; Shaw, P.A.; Lee, H.; Toh, S.W.; Connolly, J.G.; Dandreo, K.J.; Nalawade, V.; Li, J.; Hernández-Muñoz, J.J.; et al. Analyzing missingness patterns in real-world data using the SMDI toolkit: Application to a linked EHR-claims pharmacoepidemiology study. *BMC Med. Res. Methodol.* **2024**, *24*, 246. [CrossRef]
61. Syed, R.U.; Eden, R.; Makasi, T.; Chukwudi, I.; Mamudu, A.; Kamalpour, M.; Geeganage, D.K.; Sadeghianasl, S.; Leemans, S.J.J.; Goel, K.; et al. Digital Health Data Quality Issues: Systematic Review. *J. Med. Internet Res.* **2023**, *25*, e42615. [CrossRef]
62. Wu, Y.; Ren, M.; Chen, N. Semantics-driven improvements in electronic health records: A systematic mapping of semantic technologies and data quality indicators. *BMC Med. Inform. Decis. Mak.* **2025**, *25*, 146. [CrossRef]
63. Palojoki, S.; Lehtonen, L.; Vuokko, R. Semantic interoperability of electronic health records: Systematic review of alternative approaches for enhancing patient information availability. *JMIR Med. Inform.* **2024**, *12*, e53535. [CrossRef]
64. Cross, J.L.; Choma, M.A.; Onofrey, J.A. Bias in medical AI: Implications for clinical decision-making. *PLoS Digit. Health* **2024**, *3*, e0000651. [CrossRef]
65. Hanna, M.G.; Pantanowitz, L.; Jackson, B.; Palmer, O.; Visweswaran, S.; Pantanowitz, J.; Deebajah, M.; Rashidi, H.H. Ethical and Bias Considerations in Artificial Intelligence/Machine Learning. *Mod. Pathol.* **2025**, *38*, 100686. [CrossRef]
66. Pham, T. Ethical and legal considerations in healthcare AI: Innovation and policy for safe and fair use. *R. Soc. Open Sci.* **2025**, *12*, 241873. [CrossRef]
67. Razzaghi, H.; Davies, A.G.; Boss, S.; Bunnell, H.T.; Chen, Y.; Chrischilles, E.A.; Dickinson, K.; Hanauer, D.; Huang, Y.; Ilunga, K.T.S.; et al. Systematic data quality assessment of electronic health record data to evaluate study-specific fitness: Report from the PRESERVE research study. *PLoS Digit. Health* **2024**, *3*, e0000527. [CrossRef]

68. Coutinho-Almeida, J.A.; Saez, C.; Correia, R.; Rodrigues, P.P. Development and initial validation of a data quality evaluation tool in obstetrics real-world data through HL7-FHIR interoperable Bayesian networks and expert rules. *JAMIA Open* **2024**, *7*, ooae062. [CrossRef]
69. Ozonze, O.; Scott, P.J.; Hopgood, A.A. Automating Electronic Health Record Data Quality Assessment. *J. Med. Syst.* **2023**, *47*, 23. [CrossRef]
70. Gu, B.; Shao, V.; Liao, Z.; Carducci, V.; Brufau, S.R.; Yang, J. Scalable information extraction from free text electronic health records using large language models. *BMC Med. Res. Methodol.* **2025**, *25*, 23. [CrossRef] [PubMed]
71. Dahl, S.; Bøgsted, M.; Sagi, T.; Vesteghem, C. Performance of Natural Language Processing for Information Extraction From Electronic Health Records Within Cancer: Systematic Review. *JMIR Med. Inform.* **2025**, *13*, e68707. [CrossRef] [PubMed]
72. Ohlsen, T.; Ingenerf, J.; Essenwanger, A.; Drenkhahn, C. PCEtoFHIR: Decomposition of Postcoordinated SNOMED CT Expressions for Storage as HL7 FHIR Resources. *JMIR Med. Inform.* **2024**, *12*, e57853. [CrossRef] [PubMed]
73. Kho, A.N.; Yu, J.; Bryan, M.S.; Gladfelter, C.; Gordon, H.S.; Grannis, S.; Madden, M.; Mendonca, E.; Mitrovic, V.; Shah, R.; et al. Privacy-Preserving Record Linkage to Identify Fragmented Electronic Medical Records in the All of Us Research Program. In Proceedings of the Machine Learning and Knowledge Discovery in Databases, International Workshops of ECML PKDD 2019, Würzburg, Germany, 16–20 September 2019; Communications in Computer and Information Science; Cellier, P.; Driessens, K., Eds.; Springer: Cham, Switzerland, 2020; Volume 1168, pp. 79–87. [CrossRef]
74. Wani, R.U.Z.; Can, O. FED-EHR: A Privacy-Preserving Federated Learning Framework for Decentralized Healthcare Analytics. *Electronics* **2025**, *14*, 3261. [CrossRef]
75. Zhang, F.; Kreuter, D.; Chen, Y.; Dittmer, S.; Tull, S.; Shadbahr, T.; Schut, M.; Asselbergs, F.; Kar, S.; Sivapalaratnam, S.; et al. Recent methodological advances in federated learning for healthcare. *Patterns* **2024**, *5*, 101006. [CrossRef]
76. Gottlieb, M.; Palter, J.; Westrick, J.; Peksa, G.D. Effect of Medical Scribes on Throughput, Revenue, and Patient and Provider Satisfaction: A Systematic Review and Meta-analysis. *Ann. Emerg. Med.* **2021**, *77*, 180–189. [CrossRef]
77. Kalsy, M.; Burant, R.; Ball, S.; Pohnert, A.; Dolansky, M.A. A human centered design approach to define and measure documentation quality using an EHR virtual simulation. *PLoS ONE* **2024**, *19*, e0308992. [CrossRef]
78. Goebel, J.; Sidle, J.; Aspiras, A.; Fow, L.; McCann-Pineo, M.; Li, T. Reducing Physician Burnout Through Workflow Redesign: A Quality-Improvement Initiative. *Cureus* **2025**, *17*, e85799. [CrossRef]
79. Rajamani, S.; Hultman, G.; Bakker, C.; Melton, G.B. The role of organizational culture in health information technology implementations: A scoping review. *Learn. Health Syst.* **2021**, *6*, e10299. [CrossRef]
80. Campione, J.; Liu, H. Perceptions of hospital electronic health record (EHR) training, support, and patient safety by staff position and tenure. *BMC Health Serv. Res.* **2024**, *24*, 955. [CrossRef]
81. Khatiwada, P.; Yang, B.; Lin, J.C.; Blobel, B. Patient-Generated Health Data (PGHD): Understanding, Requirements, Challenges, and Existing Techniques for Data Security and Privacy. *J. Pers. Med.* **2024**, *14*, 282. [CrossRef]
82. Modi, S.; Feldman, S.S. The Value of Electronic Health Records Since the Health Information Technology for Economic and Clinical Health Act: Systematic Review. *JMIR Med. Inform.* **2022**, *10*, e37283. [CrossRef]
83. Seligson, M.T.; Lyden, S.P.; Caputo, F.J.; Kirksey, L.; Rowse, J.W.; Smolock, C.J. Improving clinical documentation of evaluation and management care and patient acuity improves reimbursement as well as quality metrics. *J. Vasc. Surg.* **2021**, *74*, 2055–2062. [CrossRef] [PubMed]
84. Boyd, A.D.; Gonzalez-Guarda, R.; Lawrence, K.; Patil, C.L.; Ezenwa, M.O.; O'Brien, E.C.; Paek, H.; Braciszewski, J.M.; Adeyemi, O.; Cuthel, A.M.; et al. Equity and bias in electronic health records data. *Contemp. Clin. Trials* **2023**, *130*, 107238. [CrossRef] [PubMed]
85. Wiley, K.; Blackburn, J.; Mendonca, E.; Menachemi, N.; De Groot, M. Associations between patient portal use and electronic health record (EHR) data timeliness in type 2 diabetes mellitus care. *J. Diabetes Metab. Disord.* **2024**, *23*, 2073–2080. [CrossRef] [PubMed]
86. Schnipper, J.L.; Gandhi, T.K.; Wald, J.S.; Grant, R.W.; Poon, E.G.; Volk, L.A.; Businger, A.; Williams, D.H.; Siteman, E.; Buckel, L.; et al. Effects of an online personal health record on medication accuracy and safety: A cluster-randomized trial. *J. Am. Med. Inform. Assoc.* **2012**, *19*, 728–734. [CrossRef]
87. Lin, J.; Bates, S.M.; Allen, L.N.; Wright, M.; Mao, L.; Kidd, M. Integrating Mobile Health App Data Into Electronic Medical or Health Record Systems and Its Impact on Health Care Delivery and Patient Health Outcomes: Scoping Review. *JMIR Mhealth Uhealth* **2025**, *13*, e66650. [CrossRef] [PubMed]
88. Wang, S.H.; Lee, Y.L.; Su, E.C.; Tsai, C.H. Role of health information technology in enhancing chronic disease management: A scoping review protocol. *BMJ Open* **2025**, *15*, e093220. [CrossRef] [PubMed]
89. Kim, Y.; Kim, G.; Cho, H.; Kim, Y.; Choi, M. Application of Patient-Generated Health Data Among Older Adults With Cancer: Scoping Review. *J. Med. Internet Res.* **2025**, *27*, e57379. [CrossRef] [PubMed]
90. Wamala Andersson, S.; Gonzalez, M.P. Digital health literacy—A key factor in realizing the value of digital transformation in healthcare. *Front. Digit. Health* **2025**, *7*, 1461342. [CrossRef]

91. Iqbal, S.; Moniz, S.; Bennin, F.; Garavito, G.A.; de Koning, R.; Yu, R.; Vindrola-Padros, C. Exploring the implementation of a data trust committee: A qualitative evaluation of processes and practices. *Res. Involv. Engagem.* **2025**, *11*, 19. [CrossRef]
92. Abdolkhani, R.; Gray, K.; Borda, A.; DeSouza, R. Recommendations for the Quality Management of Patient-Generated Health Data in Remote Patient Monitoring: Mixed Methods Study. *JMIR Mhealth Uhealth* **2023**, *11*, e35917. [CrossRef]
93. Elfouly, T.; Alouani, A. A Comprehensive Survey on Wearable Computing for Mental and Physical Health Monitoring. *Electronics* **2025**, *14*, 3443. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

systems

MDPI

Article

A Four-Party Evolutionary Game Analysis of Silver Economy Data Sharing Based on Digital Platforms

Zhiyong Zhang [1], Liyan Xia [1], Yan Shi [2] and Yongqiang Shi [1,*]

[1] Department of Electronic Business, South China University of Technology, Guangzhou 510006, China; zhyzhang@scut.edu.cn (Z.Z.); 202420159497@mail.scut.edu.cn (L.X.)
[2] School of Economics and Management, Guangxi University of Science and Technology, Liuzhou 545006, China; 2023041@gxust.edu.cn
* Correspondence: yqshi@scut.edu.cn

Abstract

As the aging society progresses, it is particularly important to strengthen the sharing of silver economy data to promote the development of the silver economy. This paper focuses on analyzing the mechanism by which digital platforms promote silver economy data sharing and constructs an evolutionary game model that includes government departments, digital platforms, enterprises, and elderly people. On this basis, the stability of the strategies of each subject in the system is analyzed, and the influence of key parameters is also discussed. The simulation draws the following conclusions. Firstly, initial strategy proportions significantly influence evolutionary directions. Higher initial proactive participation increases the probability of convergence to the optimal state. Secondly, digital platforms are driven by government regulation intensity, user complaint probabilities, and reputational losses. Increasing fines and user complaint probabilities incentivize platforms to offer high-quality protection. Thirdly, government departments can initially incentivize enterprises and elderly people to participate in data sharing through subsidies and tax incentives and build a long-term driving mechanism by improving regulatory mechanisms and enhancing digital literacy among the elderly people. The research results can serve as a reference for government departments to promote data sharing in the silver economy.

Keywords: silver economy; data sharing; evolutionary game; digital platform; pareto optimality

1. Introduction

Amid the accelerating global aging trend, the silver economy has become a pivotal strategic field for addressing demographic restructuring and reshaping elderly care ecosystems. In January 2024, China's State Council issued the *Opinions of the General Office of the State Council on Developing the Silver-Hair Economy and Enhancing the Well-Being of the Elderly*, explicitly advocating for strengthened national-level elderly care data sharing to empower industrial development through data elements [1]. This policy direction highlights the core position of data elements in the development of the silver economy. Currently, China's silver economy faces challenges such as supply–demand mismatches, service fragmentation, and insufficient industrial coordination [2]. These issues stem from entrenched data barriers among governments, enterprises, and elderly people. Therefore, promoting data sharing and unleashing the value of data are of great significance for the development of the silver economy.

Data sharing refers to individuals voluntarily revealing their information and data to others [3]. Promoting data sharing is beneficial for all parties involved in the silver economy. For government departments, promptly obtaining information on enterprises and understanding the current situation regarding elderly care demand can help accurately assess the quality and efficiency of silver economy services, identify problems, and make necessary policy adjustments. For enterprises, timely understanding of the changes in the physical and psychological conditions of the elderly people will help to effectively adjust resource allocation on the supply-side, solve supply and demand mismatch problems, and thus obtain higher profits. For elderly people, mastering more business operation data also helps them select trustworthy and excellent enterprises and enjoy higher-quality services.

Digital platforms for the silver economy are a full integration of elderly care services and information technology [4]. By coordinating and integrating resources and overcoming information asymmetry and other issues, they can effectively reduce costs, improve service quality and efficiency, and promote the high-quality development of the silver economy. Existing theoretical research predominantly focuses on the unidirectional empowerment of data elements in elderly care industries [5,6], overlooking the intermediary mechanisms linking data to supply structure optimization. While some studies explore digital technologies in elderly care applications [7–9], few address how multi-stakeholder interactions within platform ecosystems shape supply-side transformations.

Therefore, researching how to achieve data sharing through digital platforms is of enormous significance for promoting the high-quality development of the silver economy. Data sharing between government departments, digital platforms, enterprises, and elderly people is a dynamic and complex process, and changes in the strategies of all parties may have a profound impact on results. We construct a model from the perspective of the government to explore specific pathways for achieving silver economy data sharing under government guidance. To enhance the practical significance of this research, numerical simulation experiments are employed to identify key factors affecting data sharing and to propose actionable policy proposals to help government departments better guide silver economy data sharing.

The article is structured as follows. Section 2 conducts a literature review, analyzing the concept and development history of the silver economy, as well as current problems and solutions. In addition, the current application status of evolutionary game theory (EGT) in the fields of data sharing and elderly care is summarized, providing a reference for the application of EGT to analyze silver economy data sharing. In Section 3, we analyze the strategy stability of government departments, digital platforms, enterprises, and elderly people. In Section 4, we verify the accuracy of the theoretical analysis results through numerical simulation experiments and analyze the impact of various parameters on the evolutionary process of silver economy data sharing. In Section 5, we summarize the conclusions drawn through simulation, with a focus on recommendations for government departments.

2. Literature Review

2.1. The Development of the Silver Economy

The concept of the "silver economy" was first proposed in the 1970s in Japan, specifically referring to industries providing diversified products and services tailored to the needs of the elderly [10]. In 2008, as advanced global economies entered aging societies, a senior-oriented market encompassing consumption, healthcare, and other fields emerged, prompting the European Parliament to formally adopt the term "silver economy" [11]. In China, the silver economy is defined as "the sum of economic activities that provide products or services to the elderly, as well as preparations for the aging stage" [1].

https://doi.org/10.3390/systems14010027

Guided by the "current status—problems—recommendations" research framework, Chinese scholars have identified key challenges in the silver economy, including insufficient market penetration, frequent fraud incidents, and irrational supply structures [12]. Strategic proposals emphasize the coordinated development of population and high-quality economic linkages [13]. Research hotspots focus on the growth of silver consumption markets [14], comparative studies of silver economy models in developed countries [15], and the convergence of digital and silver economies [6]. In addition, partial studies dissect the silver economy from managerial perspectives [16] and propose suggestions such as establishing a "comprehensive coordinated fiscal and taxation system" [17] and a KPI assessment mechanism [18].

Domestic research predominantly employs qualitative, empirical, and quantitative methods to examine issues and pathways for high-quality development in the silver economy. However, most studies overlook the critical role of data as a production factor in driving this sector. In contrast, some international literature highlights data's transformative potential. He et al., using regression models, analyzed how the digital economy influences consumption upgrades among China's elderly, concluding that digital advancement significantly expands seniors' economic participation and material well-being [19]. Guan et al. applied case studies, cluster analysis, and Bayesian models to explore how digital transformation in health and elderly care industries enhances service delivery [20]. Fan et al., through OLS regression, demonstrated that digital financial inclusion (DFI) boosts elderly consumption expenditure, advocating improved digital financial services to stimulate silver economy growth [21]. Zeng et al., utilizing descriptive statistics and cluster analysis, identified artificial intelligence and digital technologies as catalysts for innovation in digital healthcare to meet silver economy demands [22]. Ling et al. proposed a data-driven framework to optimize the distribution of elderly care facilities [23].

In summary, current research findings indicate that data elements have a promoting effect on the development of the silver economy, but there is a lack of research on how to tap their potential value. Therefore, applying evolutionary game theory to analyze strategic interactions among multiple entities in the silver economy data market will help explore how to maximize the value of data elements and promote the high-quality development of the silver economy.

2.2. Evolutionary Game Theory

Game theory provides a robust mathematical framework for analyzing and predicting the strategic behaviors of participants [24], significantly enhancing the decision-making capabilities of stakeholders within game systems [25]. In complex social decision-making scenarios, participants often exhibit bounded rationality and make suboptimal choices due to incomplete information [26]. Evolutionary game theory (EGT), which emphasizes dynamic equilibria [27] and does not require perfect rationality or complete information transparency [28], is particularly suited to studying such intricate societal issues.

Recently, EGT has been increasingly applied to explore challenges in the silver economy and elderly care services. Shi et al. developed a tripartite evolutionary game model to analyze the behavioral strategies of governments, smart elderly care service providers, and elderly users [29]. Wang et al. constructed a tripartite model involving local governments, private elderly care institutions, and the public to inform regulatory policy design [30]. Li et al. investigated the impact of public participation on governmental oversight in public–private partnership projects using EGT [24]. Zhang et al. focused on strategy interactions among governments, elderly people, and private elderly care institutions to guide institutional regulatory policies [31]. Wang et al. examined dynamic interactions among governments, elderly care service providers, platforms, and elderly users [32]. These stud-

ies collectively demonstrate the feasibility of applying EGT to address challenges in the silver economy and elderly care.

In addition, EGT has also been widely applied in data sharing. Zhang et al. focused on strategic interactions among governments, health service centers, and hospitals, analyzing the impact of patient evaluation, government punishment, and rewards on knowledge sharing among medical institutions [33]. Zhang et al. constructed an evolutionary game model including government, data centers, data demand medical institutions, and data supply medical institutions, and conclude that strong regulation, subsidies, and reward and punishment mechanisms are all conducive to promoting medical data sharing [34]. Li et al. established a model between universities to identify the difficulties in sharing resources among universities [35]. Dong et al. studied the issue of data sharing between enterprises under government supervision, exploring the impact mechanism of factors such as government penalties, rewards, and compensation fees [36].

In summary, although some scholars use EGT to study issues in the silver economy, they mainly focus on the quality supervision of elderly care services, and there is still little research exploring the game theory relationship among various entities in silver economy data sharing. To address this gap, this study establishes a model involving government departments, digital platforms, enterprises, and elderly people, systematically studying their strategic decisions in different scenarios.

2.3. Mechanism of Digital Platforms in the Silver Economy

Digital platforms are different from traditional platforms, which are limited by region and cost. With help from the Internet and digital technologies, they can promote the development of the silver economy [37]. The emergence of digital platforms has disrupted traditional elderly care service and injected new vitality into the development of the silver economy. Currently, scholars both domestically and internationally have conducted extensive research on the application of digital platforms in the silver economy. It mainly includes data-driven smart elderly care service content and modes [38], improvements in elderly care service quality driven by digital platforms [39], the construction of information platforms to achieve information sharing between communities and governments [40], and the integrated development of the digital economy and silver economy [6].

Existing research has shown the positive role of digital platforms in the development of the silver economy. The role of digital platforms is to drive data-driven optimization of the supply-side structure of the silver economy, achieving precise matching between supply and demand. Although some scholars have noticed these trends, in-depth research on the mechanisms of digital platforms have not been conducted. Therefore, this article analyzes the strategic interaction logic among government departments, digital platforms, enterprises, and elderly people, and elaborates on the mechanism of achieving data sharing in the silver economy through digital platforms, providing references for promoting the development of the silver economy.

3. Model

3.1. Problem Description

In the digital era, data has become a critical strategic asset for enterprises [41]. It plays an increasingly important role in modern economies [42]. During the development of the silver economy, data sharing presents a duality of benefits and challenges. On the one hand, data sharing helps elderly people better understand the operational status of enterprises and changes in government policies. At the same time, it also helps enterprises grasp changes in user needs and adjust resource allocation. Furthermore, the government can understand the development of the silver economy industry, identify problems, and make

policy adjustments. On the other hand, it incurs certain costs and risks, including explicit costs of data sharing and implicit costs caused by data leakage risks. From the government's perspective, it is crucial to support silver economy data sharing. The government's incentive and guidance role, as well as regulatory mechanisms, help facilitate the orderly sharing of data in the silver economy, thereby promoting the development of the silver economy industry, improving social benefits, and enhancing the living standards of the elderly.

The silver economy digital platform represents the integration of the silver economy with new-generation information technologies, shaping a new model of data sharing dominated by digital platforms. This approach enhances the efficiency, quality, and structure of elderly care service provision, ultimately forming a "data collection–information processing–autonomous decision-making" supply model. Thus, the digital platform-driven silver economy establishes interconnectedness among government departments, digital platforms, enterprises, and elderly people. Figure 1 shows the service supply model of the silver economy centered on digital platforms.

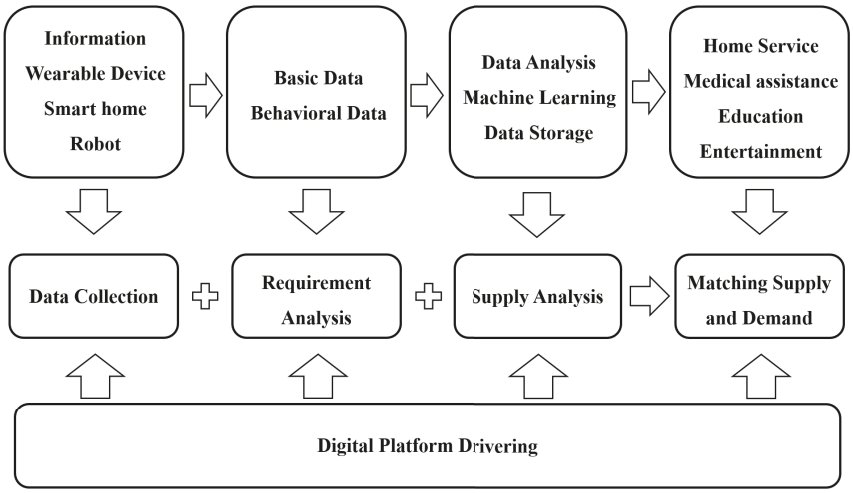


Figure 1. Digital platform-driven silver economy service supply model.

Supply is key to elderly care services in the silver economy, and the crux of achieving precise supply–demand matching lies in using data to accurately analyze elderly users' needs. Digital platforms can transform fragmented elderly care services into integrated supply by uniting government departments, enterprises, community service organizations, and others to deliver continuous, coherent, and real-time services to elderly users. Data sharing optimizes the supply structure of silver economy services, improving supply efficiency and service quality. In this process, the task of digital platforms is to maintain the normal operation of data sharing, ensure data security for all parties through high-tech means, and prevent security incidents such as privacy breaches (Figure 2).

This study constructs a government-involved evolutionary game model for silver economy data sharing. The process of silver economy data sharing under government guidance includes four main entities: government departments, digital platforms, enterprises, and elderly people. Figure 3 below illustrates the logical relationships among the four-party game participants. Nevertheless, this model does have inherent limitations. To make the model analyzable, the paper simplifies the assumptions regarding four-party behavior and decision-making, which may be disconnected from real behavior.

As the guiding party, the strategy set of government departments is {G_1 positive promotion, G_2 passive promotion}. Positive promotion means that government departments carry out mandatory promotion of silver economy data sharing, establish sound supervision and punishment mechanisms, strictly monitor digital platform behavior, and

issue relevant policies to encourage enterprises and elderly people to participate in data sharing through subsidies, tax reductions, and other means. Passive promotion means that government departments simplify processes to reduce administrative costs and allow the market to develop freely. As intermediaries for data sharing, the strategy set of digital platforms is {D_1, high-quality protection, D_2, low-quality protection}. When choosing to provide high-quality protection, digital platforms will incur additional costs to inspect vulnerabilities in data collection and transmission processes, reduce the probability of data leakage, and protect user data security. As providers and demanders of silver economy data, the strategy sets of enterprises and elderly people are as follows: {E_1 cooperation, E_2 non-cooperation} and {P_1 sharing data, P_2 non-sharing data}. Cooperation refers to enterprises sharing some business data with digital platforms and obtaining necessary policy information and elderly-related data from the platforms. Sharing data refers to elderly people sharing personal data with platforms and enterprises, allowing businesses in different regions to read and analyze it and enjoy personalized services from enterprises.

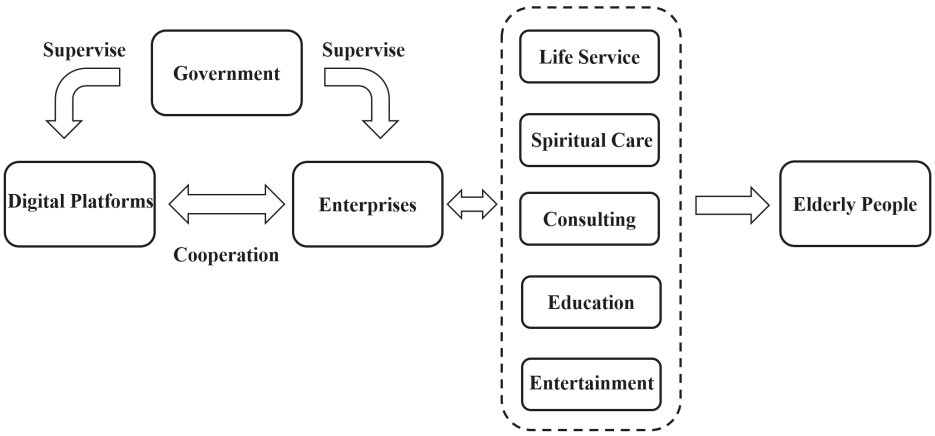


Figure 2. Digital platforms drive the supply-side structural optimization.

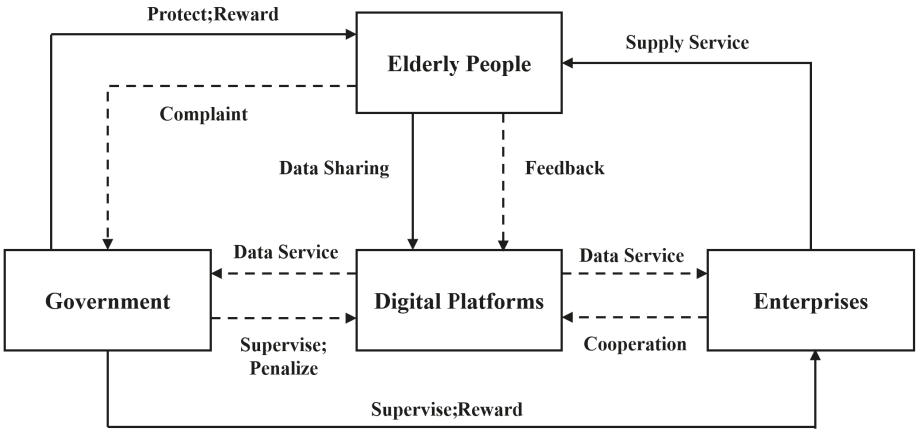


Figure 3. Relationships among government departments, digital platforms, enterprises, and elderly people.

3.2. Model Assumptions

The model in this article is mainly based on the following assumptions.

Hypothesis 1. *Participants include government departments (G), digital platforms (D), enterprises (E), and elderly people (P). All parties exhibit bounded rationality. Each agent probabilistically selects optimal strategies to maximize its respective utility. The probabilities of government departments enforcing positive promotion (G_1) or passive promotion (G_2) are (x, $1-x$). The*

probabilities of digital platforms offering high-quality protection (D_1) or low-quality (D_2) protection are ($y,\ 1-y$). The probabilities of enterprises performing cooperation (E_1) or non-cooperation (E_2) are ($z,\ 1-z$). The probabilities that elderly people choose sharing data (P_1) or non-sharing data (P_2) with digital platforms are ($p,\ 1-p$). Here $x, y, z, p \in [0,\ 1]$.

Hypothesis 2. *The government's costs for G_1 and G_2 strategy are C_{Gh} and C_{Gl}, respectively. When enterprises choose E_1 strategy and elderly people choose P_1 strategy, a positive impact B_G is generated for silver economy data sharing.*

Hypothesis 3. *The digital platforms earn R_D by providing data services to government departments. If enterprises select E_1 strategy, platforms gain additional revenue ΔR_D. Costs for D_1 and D_2 strategies are C_{Dh} and C_{Dl} ($C_{Dh} > C_{Dl}$), respectively. Modern threats such as cyberattacks can easily cause security incidents such as data breaches, and digital platforms can reduce the probability of data breaches by providing high-quality protection. Data leakage probabilities under D_1 and D_2 are α and $\beta(\alpha < \beta)$, respectively.*

Hypothesis 4. *Enterprises incur a cost C_E for cooperation (E_1) (including lost exclusivity and channel maintenance). Cooperation enables enterprises to timely understand government policy information and industry trends from digital platforms, achieve resource optimization and innovation through data analysis, and generate profits R_E. When government departments enforce positive promotion, enterprises receive tax incentives B_E. If elderly people share data with digital platforms, enterprises gain extra revenue ΔR_E via platform recommendations. Non-cooperation (E_2) yields zero cost and zero revenue.*

Hypothesis 5. *Elderly people gain R_P from sharing data (personalized services) and subsidies B_P under government G_1 strategy. Their cost C_P include learning and time costs. Non-sharing yields zero cost and zero revenue.*

Hypothesis 6. *Losses caused by data breaches to government departments, digital platforms, enterprises, and elderly people are L_G, L_D, L_E, and L_P, respectively. When digital platforms choose D_2 strategy, the probability of elderly people discovering security issues is θ. Active government response boosts credibility (gain R_G) and penalizes platforms (reputation loss S).*

Hypothesis 7. *When digital platforms choose D_2 strategy, and government departments choose G_1 or G_2 strategy, the probabilities of discovering security risks are φ and ω, respectively. When the government discovers that digital platforms adopt low-quality protection strategies that lead to security vulnerabilities, fines ΔP will be imposed on digital platforms. When a data breach occurs, in order to encourage platforms to provide high-quality protection for user data, a fine of $(1-\lambda)P$ will be imposed on platforms that choose D_1 strategy, and a fine of $(1+\lambda)P$ will be imposed on platforms that choose D_2 strategy.*

3.3. Stability Analysis of Players' Strategy Choices

Based on Hypotheses 1–7, the profits from the diverse strategies selected by government departments, digital platforms, enterprises, and elderly people are delineated in Table 1.

Table 1. The payoff matrix.

Government Departments	Digital Platforms	Enterprises	Elderly People	
			Sharing Data (p)	**Non-Sharing Data ($1-p$)**
positive promotion (x)	high-quality protection (y)	cooperation (z)	$B_G + \alpha[(1-\lambda)P - L_G] - C_{Gh}$, $R_D + \Delta R_D - C_{Dh} - \alpha[(1-\lambda)P + L_D]$, $R_E + B_E + \Delta R_E - C_E - \alpha L_E$, $R_P + B_P - C_P - \alpha L_P$	$\alpha[(1-\lambda)P - L_G] - C_{Gh}$, $R_D + \Delta R_D - C_{Dh} - \alpha[(1-\lambda)P + L_D]$, $R_E + B_E - C_E - \alpha L_E$, 0
		non-cooperation ($1-z$)	$\alpha[(1-\lambda)P - L_G] - C_{Gh}$, $R_D - C_{Dh} - \alpha[(1-\lambda)P + L_D]$, 0, $R_P + B_P - C_P - \alpha L_P$	$\alpha[(1-\lambda)P - L_G] - C_{Gh}$, $R_D - C_{Dh} - \alpha[(1-\lambda)P + L_D]$, 0, 0
	low-quality protection ($1-y$)	cooperation (z)	$B_G + \beta[(1+\lambda)P - L_G] - C_{Gh} + \varphi\Delta P + \theta R_G$, $R_D + \Delta R_D - C_{Dl} - \beta[(1+\lambda)P + L_D] - \theta S - \varphi\Delta P$, $R_E + B_E + \Delta R_E - C_E - \beta L_E$, $R_P + B_P - C_P - \beta L_P$	$\beta[(1+\lambda)P - L_G] - C_{Gh} + \varphi\Delta P$, $R_D + \Delta R_D - C_{Dl} - \beta[(1+\lambda)P + L_D] - \varphi\Delta P$, $R_E + B_E - C_E - \beta L_E$, 0
		non-cooperation ($1-z$)	$\beta[(1+\lambda)P - L_G] - C_{Gh} + \varphi\Delta P$, $R_D - C_{Dl} - \beta[(1+\lambda)P + L_D] - \theta S - \varphi\Delta P$, 0, $R_P + B_P - C_P - \beta L_P$	$\beta[(1+\lambda)P - L_G] - C_{Gh} + \varphi\Delta P$, $R_D - C_{Dl} - \beta[(1+\lambda)P + L_D] - \varphi\Delta P$, 0, 0
passive promotion ($1-x$)	high-quality protection (y)	cooperation (z)	$B_G + \alpha[(1-\lambda)P - L_G] - C_{Gl}$, $R_D + \Delta R_D - C_{Dh} - \alpha[(1-\lambda)P + L_D]$, $R_E + \Delta R_E - C_E - \alpha L_E$, $R_P - C_P - \alpha L_P$	$\alpha[(1-\lambda)P - L_G] - C_{Gl}$, $R_D + \Delta R_D - C_{Dh} - \alpha[(1-\lambda)P + L_D]$, $R_E - C_E - \alpha L_E$, 0
		non-cooperation ($1-z$)	$\alpha[(1-\lambda)P - L_G] - C_{Gl}$, $R_D - C_{Dh} - \alpha[(1-\lambda)P + L_D]$, 0, $R_P - C_P - \alpha L_P$	$\alpha[(1-\lambda)P - L_G] - C_{Gl}$, $R_D - C_{Dh} - \alpha[(1-\lambda)P + L_D]$, 0, 0
	low-quality protection ($1-y$)	cooperation (z)	$B_G + \beta[(1+\lambda)P - L_G] - C_{Gl} + \omega\Delta P$, $R_D + \Delta R_D - C_{Dl} - \beta[(1+\lambda)P + L_D] - \theta S - \omega\Delta P$, $R_E + \Delta R_E - C_E - \beta L_E$, $R_P - C_P - \beta L_P$	$\beta[(1+\lambda)P - L_G] - C_{Gl} + \omega\Delta P$, $R_D + \Delta R_D - C_{Dl} - \beta[(1+\lambda)P + L_D] - \omega\Delta P$, $R_E - C_E - \beta L_E$, 0
		non-cooperation ($1-z$)	$\beta[(1+\lambda)P - L_G] - C_{Gl} + \omega\Delta P$, $R_D - C_{Dl} - \beta[(1+\lambda)P + L_D] - \theta S - \omega\Delta P$, 0, $R_P - C_P - \beta L_P$	$\beta[(1+\lambda)P - L_G] - C_{Gl} + \omega\Delta P$, $R_D - C_{Dl} - \beta[(1+\lambda)P + L_D] - \omega\Delta P$, 0, 0

In the absence of complete information, each player constantly changes their strategy to adapt to the activities of other players [43]. Therefore, after obtaining the payoff matrix, the replication dynamic equations can be derived by calculating the expected profit of each player [44], so as to analyze the stability of the strategies of government departments, digital platforms, enterprises, and elderly people [45].

3.3.1. Stability Analysis for Government Departments

The expected revenue of government departments that select G_1 strategy (G_2) is $M_{G1}(M_{G2})$, and the average expected revenue is M_G.

$$M_{G1} = zpB_G + \alpha y[(1-\lambda)P - L_G] + \beta(1-y)[(1+\lambda)P - L_G] + \varphi(1-y)\Delta P + (1-y)p\theta R_G - C_{Gh} \tag{1}$$

$$M_{G2} = zpB_G + \alpha y[(1-\lambda)P - L_G] + \beta(1-y)[(1+\lambda)P - L_G] + \omega(1-y)\Delta P - C_{Gl} \tag{2}$$

$$\begin{aligned} M_G = zpB_G + \alpha y[(1-\lambda)P - L_G] + \beta(1-y)[(1+\lambda)P - L_G] + x\varphi(1-y)\Delta P + x(1-y)p\theta R_G \\ +(1-x)\omega(1-y)\Delta P - xC_{Gh} - (1-x)C_{Gl} \end{aligned} \tag{3}$$

By using the Malthusian equation [46], the replicator dynamic equation and its first-order derivative are as follows:

$$\begin{aligned} F(x) &= x(M_{G1} - M_G) \\ &= x(1-x)[-C_{Gh} + C_{Gl} + (\varphi - \omega)(1-y)\Delta P + (1-y)p\theta R_G] \end{aligned} \tag{4}$$

$$F'(x) = \frac{d(F(x))}{dx} = (1-2x)[-C_{Gh} + C_{Gl} + (\varphi - \omega)(1-y)\Delta P + (1-y)p\theta R_G] \tag{5}$$

By using the stability theorem of differential equations [47–49], the probability of government departments decision-making in a stable state must satisfy $F(x) = 0$ and $F'(x) < 0$.

Proposition 1. *When $p < p^*$, government departments choose the passive promotion strategy. When $p > p^*$, if $y > y^*$, government departments choose the passive promotion strategy; if $y < y^*$, government departments choose the positive promotion strategy. Here, $y^*$and p^* are as follows:*

$$y^* = \frac{-C_{Gh} + C_{Gl} + (\varphi - \omega)\Delta P + p\theta R_G}{(\varphi - \omega)\Delta P + p\theta R_G},\ p^* = \frac{C_{Gh} - C_{Gl} - (\varphi - \omega)\Delta P}{\theta R_G}$$

Proof of Proposition 1. Let $G(y) = -C_{Gh} + C_{Gl} + (\varphi - \omega)(1-y)\Delta P + (1-y)p\theta R_G$, because $\frac{\partial G(y)}{\partial y} = -(\varphi - \omega)\Delta P - p\theta R_G < 0$, $G(y)$ is a decreasing function of y. Next, we discuss this proposition under two situations.

Case 1: When $C_{Gh} - C_{Gl} - (\varphi - \omega)\Delta P < 0$, $p^* < 0$. For $p \in [0,1]$, the probability of $p > p^*$ is 1, then $y^* > 0$. If $y > y^*$, $G(y) < G(y^*) = 0$, $F'(x)|_{x=0} = G(y) < 0$, $F'(x)|_{x=1} = -G(y) > 0$, thus $x = 0$ is the stable strategy. If $y < y^*$, $G(y) > G(y^*) = 0$, $F'(x)|_{x=0} = G(y) > 0$, $F'(x)|_{x=1} = -G(y) < 0$, meaning that $x = 1$ is the stable strategy. If $y = y^*$, $G(y) = G(y^*) = 0$, $F'(x)|_{x=0} = F'(x)|_{x=1} = 0$, government departments cannot determine a stable strategy.

Case 2: When $C_{Gh} - C_{Gl} - (\varphi - \omega)\Delta P > 0$, $p^* > 0$. When $p > p^*$, $y^* > 0$. If $y > y^*$, $G(y) < G(y^*) = 0$, $F'(x)|_{x=0} = G(y) < 0$, $F'(x)|_{x=1} = -G(y) > 0$, thus $x = 0$ is the stable strategy. If $y < y^*$, $G(y) > G(y^*) = 0$, $F'(x)|_{x=0} = G(y) > 0$, $F'(x)|_{x=1} = -G(y) < 0$, meaning that $x = 1$ is the stable strategy. If $y = y^*$, $G(y) = G(y^*) = 0$, $F'(x)|_{x=0} = F'(x)|_{x=1} = 0$, government departments cannot determine a stable strategy. When $p < p^*$, $y^* < 0$, for $y \in [0,1]$, the probability of $y > y^*$ is 1. Therefore, when $y \in [0,1]$, $G(y) < G(y^*) = 0$, $F'(x)|_{x=0} = G(y) < 0$, $F'(x)|_{x=1} = -G(y) > 0$, thus $x = 0$ is the stable strategy. □

Proposition 1 suggests that regulatory measures implemented by government departments toward digital platforms influence the stable strategies of those platforms, and in turn, the strategic choices made by these platforms affect the government's strategic decisions. When the government intensifies regulatory oversight and imposes higher penalties, the threshold y^* decreases. Consequently, when government departments enforce positive promotion and enhance regulatory intensity, digital platforms—facing heavier penalties—opt for the high-quality protection strategy. When digital platforms choose to provide high-quality protection, the government's stable strategy evolves from positive promotion to passive promotion.

Based on the above analysis, the phase diagram of government departments is shown in Figure 4.

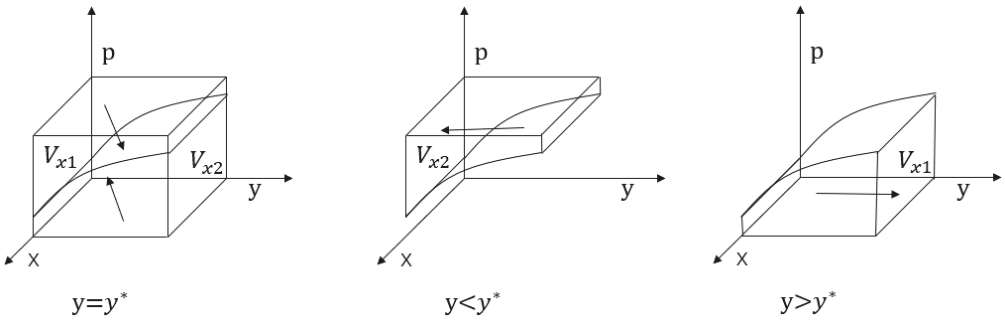


Figure 4. The phase diagram of government departments.

Based on the above analysis, we solve for V_{x1} and V_{x2}:

$$V_{x1} = \int_0^1 \int_0^1 y^* dydx = \frac{-C_{Gh} + C_{Gl} + (\varphi - \omega)\Delta P + p\theta R_G}{(\varphi - \omega)\Delta P + p\theta R_G} = 1 - \frac{C_{Gh} - C_{Gl}}{(\varphi - \omega)\Delta P + p\theta R_G} \tag{6}$$

$$V_{x2} = 1 - V_{x1} = \frac{C_{Gh} - C_{Gl}}{(\varphi - \omega)\Delta P + p\theta R_G} \tag{7}$$

Corollary 1. *The probability of government departments choosing the positive promotion strategy is positively correlated with elderly people's complaint probability, reputation benefits obtained from handing complaints, fines, and the cost of choosing the passive promotion strategy, and negatively correlated with the costs of choosing the positive promotion strategy.*

Proof. V_{x1} and V_{x2} represent the probabilities of government departments choosing the positive promotion strategy and the passive promotion strategy, respectively. Calculate the partial derivative V_{x1} with respect to θ, R_G, ΔP, C_{Gl}, C_{Gh}: $\frac{\partial V_{x1}}{\partial \theta} = \frac{(C_{Gh} - C_{Gl})pR_G}{((\varphi-\omega)\Delta P + p\theta R_G)^2} > 0$, $\frac{\partial V_{x1}}{\partial R_G} = \frac{(C_{Gh} - C_{Gl})p\theta}{((\varphi-\omega)\Delta P + p\theta R_G)^2} > 0$, $\frac{\partial V_{x1}}{\partial \Delta P} = \frac{(C_{Gh} - C_{Gl})(\varphi-\omega)}{((\varphi-\omega)\Delta P + p\theta R_G)^2} > 0$, $\frac{\partial V_{x1}}{\partial C_{Gl}} = \frac{1}{(\varphi-\omega)\Delta P + p\theta R_G} > 0$, $\frac{\partial V_{x1}}{\partial C_{Gh}} = \frac{-1}{(\varphi-\omega)\Delta P + p\theta R_G} < 0$. □

Government departments are more likely to choose positive promotion when its cost is lower or when the cost of passive promotion gradually increases. In addition, the government's fines on digital platforms and reputation enhancement brought about by actively handling complaints from elderly people will also affect the government's strategic choices. Therefore, when digital platforms passively protect user data, leading to data leakage accidents and continuous user complaints, the government should take active measures to regulate the behavior of digital platforms and ensure the smooth progress of data sharing.

3.3.2. Stability Analysis for Digital Platforms

The expected revenue for digital platforms choosing D_1 (D_2) strategy is $M_{D1}(M_{D2})$, and the average expected revenue is M_D.

$$M_{D1} = R_D + z\Delta R_D - C_{Dh} - \alpha[(1-\lambda)P + L_D] \tag{8}$$

$$M_{D2} = R_D + z\Delta R_D - C_{Dl} - \beta[(1+\lambda)P + L_D] - [x\varphi + (1-x)\omega]\Delta P - p\theta S \tag{9}$$

$$\begin{aligned} M_D = R_D + z\Delta R_D - & y\{C_{Dh} + \alpha[(1-\lambda)P + L_D]\} \\ & -(1-y)\{C_{Dl} + \beta[(1+\lambda)P + L_D] + [x\varphi + (1-x)\omega]\Delta P + p\theta S\} \end{aligned} \tag{10}$$

The replicator dynamic equation and its first-order derivative of digital platforms are as follows:

$$F(y) = y(M_{D1} - M_D) \\ = y(1-y)\{-C_{Dh} - \alpha[(1-\lambda)P + L_D] + C_{Dl} + \beta[(1+\lambda)P + L_D] + [x\varphi + (1-x)\omega]\Delta P + p\theta S\} \tag{11}$$

$$F'(y) = \frac{d(F(y))}{dy} = (1-2y)\{-C_{Dh} - \alpha[(1-\lambda)P + L_D] + C_{Dl} + \beta[(1+\lambda)P + L_D] + [x\varphi + (1-x)\omega]\Delta P + p\theta S\} \tag{12}$$

The probability of digital platforms' decision-making in a stable state must satisfy $F(y) = 0$ and $F'(y) < 0$.

Proposition 2. *When $p > p^*$, digital platforms choose to provide high-quality protection. When $p < p^*$, if $x > x^*$, digital platforms choose to provide high-quality protection. If $x < x^*$, digital platforms choose to provide low-quality protection. Here, x^* and p^* are as follows:*

$$x^* = \frac{C_{Dh} - C_{Dl} + \alpha[(1-\lambda)P + L_D] - \beta[(1+\lambda)P + L_D] - \omega\Delta P - p\theta S}{(\varphi - \omega)\Delta P}$$

$$p^* = \frac{C_{Dh} - C_{Dl} + \alpha[(1-\lambda)P + L_D] - \beta[(1+\lambda)P + L_D] - \omega\Delta P}{\theta S}$$

Proof. Let $G(x) = -C_{Dh} - \alpha[(1-\lambda)P + L_D] + C_{Dl} + \beta[(1+\lambda)P + L_D] + [x\varphi + (1-x)\omega]\Delta P + p\theta S$. Since $\frac{\partial G(x)}{\partial x} = (\varphi - \omega)\Delta P > 0$, $G(x)$ is an increasing function with regard to x. Next, we discuss this proposition in two situations.

Case 1: When $C_{Dh} - C_{Dl} + \alpha[(1-\lambda)P + L_D] - \beta[(1+\lambda)P + L_D] - \omega\Delta P < 0$, $p^* < 0$. For $p \in [0,1]$, the probability of $p > p^*$ is 1, $x^* < 0$. Therefore, when $x \in [0,1]$, $G(x) > G(x^*) = 0$, $F'(y)|_{y=0} = G(x) > 0$, $F'(y)|_{y=1} = -G(x) < 0$, meaning that $y = 1$ is the stable strategy.

Case 2: When $C_{Dh} - C_{Dl} + \alpha[(1-\lambda)P + L_D] - \beta[(1+\lambda)P + L_D] - \omega\Delta P > 0$, $p^* > 0$. If $p > p^*$, $x^* < 0$. Therefore, when $x \in [0,1]$, $G(x) > G(x^*) = 0$, $F(y)|_{y=0} = F(y)|_{y=1} = 0, F'(y)|_{y=0} = G(x) > 0$, $F'(y)|_{y=1} = -G(x) < 0$, meaning that $y = 1$ is the stable strategy. If $p < p^*$, $x^* > 0$. When $x > x^*$, $G(x) > 0$, $F'(y)|_{y=0} = G(x) > 0$, $F'(y)|_{y=1} = -G(x) < 0$, thus $y = 1$ is the stable strategy. When $x < x^*$, $G(x) < 0$, $F'(y)|_{y=0} = G(x) < 0$, $F'(y)|_{y=1} = -G(x) > 0$, thus $y = 0$ is the stable strategy. When $x = x^*$, $G(x) = 0$, $F'(y)|_{y=0} = F'(y)|_{y=1} = 0$, meaning that digital platforms cannot determine the stable strategy.

Proposition 2 implies that government departments and elderly people influence the strategic choices of digital platforms. As more elderly people choose to share data with digital platforms, the probability of data leakage increases, and the reputational damage caused by security incidents becomes more severe. Therefore, digital platforms have to provide high-quality protection to avoid greater losses. At the same time, active regulation by government departments will also force digital platforms to choose high-quality protection strategies to avoid losses from substantial fines.

The phase diagram of digital platforms is shown in Figure 5.

Based on the above analysis, we solve for V_{y1} and V_{y2}:

$$V_{y2} = \int_0^1 \int_0^1 x^* dp dy = \frac{2\{C_{Dh} - C_{Dl} + \alpha[(1-\lambda)P + L_D] - \beta[(1+\lambda)P + L_D] - \omega\Delta P\} - p\theta S}{2(\varphi - \omega)\Delta P} \tag{13}$$

$$V_{y1} = 1 - V_{y2} = \frac{p\theta S + 2\{\varphi\Delta P + C_{Dl} + \beta[(1+\lambda)P + L_D] - C_{Dh} - \alpha[(1-\lambda)P + L_D]\}}{2(\varphi - \omega)\Delta P} \tag{14}$$

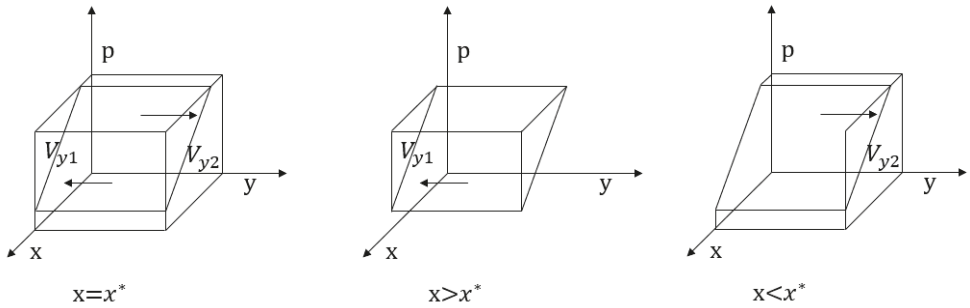


Figure 5. The phase diagram of digital platforms.

Corollary 2. *The probability of digital platforms choosing the high-quality protection strategy is positively correlated with elderly people's complaint probability, reputation losses, fines, penalty coefficients, and data leakage losses, and negatively correlated with the cost of providing high-quality protection.*

Proof. V_{y1} and V_{y2} represent the probabilities of digital platforms providing high-quality protection and low-quality protection, respectively. Calculate the partial derivative V_{y1} with respect to θ, S, P, ΔP, λ, L_D, C_{Dh}: $\frac{\partial V_{y1}}{\partial \theta} = \frac{pS}{2(\varphi-\omega)\Delta P} > 0$, $\frac{\partial V_{y1}}{\partial S} = \frac{p\theta}{2(\varphi-\omega)\Delta P} > 0$, $\frac{\partial V_{y1}}{\partial \Delta P} = \frac{2\{C_{Dh}-C_{Dl}+\alpha[(1-\lambda)P+L_D]-\beta[(1+\lambda)P+L_D]\}-p\theta S}{\Delta P^2} > 0$, $\frac{\partial V_{y1}}{\partial P} = \frac{[\beta(1+\lambda)-\alpha(1-\lambda)]}{(\varphi-\omega)\Delta P} > 0$, $\frac{\partial V_{y1}}{\partial \lambda} = \frac{(\beta-\alpha)P}{2(\varphi-\omega)\Delta P} > 0$, $\frac{\partial V_{y1}}{\partial L_D} = \frac{(\beta-\alpha)}{2(\varphi-\omega)\Delta P} > 0$, $\frac{\partial V_{y1}}{\partial C_{Dh}} = \frac{-1}{(\varphi-\omega)\Delta P} < 0$. □

Various factors influence the strategic choices of digital platforms. Increasing government penalties, increasing the probability of complaints from elderly people, and helping platforms reduce costs can encourage digital platforms to provide high-quality protection. Therefore, government regulatory agencies should establish sound regulatory systems to effectively prevent irresponsible behavior by digital platforms that provide low-quality protection. At the same time, publicity departments should increase its publicity efforts to enhance elderly people's awareness of and ability to protect data. However, if government regulatory agencies adopt overly strict regulatory measures, the probability of over-regulation increases, which may cause digital platforms to refuse participation in data sharing.

3.3.3. Stability Analysis for Enterprises

The expected revenue for enterprises choosing E_1 (E_2) strategy is $M_{E1}(M_{E2})$, and the average expected revenue is M_E.

$$M_{E1} = R_E + xB_E + p\Delta R_E - C_E - [\alpha y + \beta(1-y)]L_E \tag{15}$$

$$M_{E2} = 0 \tag{16}$$

$$M_E = z\{R_E + xB_E + p\Delta R_E - C_E - [\alpha y + \beta(1-y)]L_E\} \tag{17}$$

The replicator dynamic equation and its first-order derivative are as follows:

$$F(z) = z(M_{E1} - M_E) = z(1-z)\{R_E + xB_E + p\Delta R_E - C_E - [\alpha y + \beta(1-y)]L_E\} \tag{18}$$

$$F'(z) = \frac{d(F(z))}{dz} = (1-2z)\{R_E + xB_E + p\Delta R_E - C_E - [\alpha y + \beta(1-y)]L_E\} \tag{19}$$

The probability of enterprises' decision-making in a stable state must satisfy $F(z) = 0$ and $F'(z) < 0$.

Proposition 3. *When* $p > p^*$ *, enterprises enforce cooperation. When* $p < p^*$*, if* $x > x^*$*, enterprises enforce cooperation. If* $x < x^*$*, when* $y > y^*$*, enterprises enforce cooperation. When* $y < y^*$*, enterprises choose the non-cooperation strategy. Here,* y^**,* x^**, and* p^* *are as follows:*

$$y^* = \frac{C_E + \beta L_E - R_E - xB_E - p\Delta R_E}{(\beta - \alpha)L_E},\ x^* = \frac{C_E + \beta L_E - R_E - p\Delta R_E}{B_E},\ p^* = \frac{C_E + \beta L_E - R_E}{\Delta R_E}$$

Proof. Let $G(y) = R_E + xB_E + p\Delta R_E - C_E - [\alpha y + \beta(1-y)]L_E$, because $\frac{\partial G(y)}{\partial y} = -(\alpha - \beta)L_E > 0$, $G(y)$ is an increasing function of y. Next, we discuss this proposition in two situations.

Case 1: When $C_E + \beta L_E - R_E < 0$, $p^* < 0$. For $p \in [0,1]$, the probability of $p > p^*$ is 1, then $x^* < 0$. For $x \in [0,1]$, the probability of $x > x^*$ is 1. Therefore, $y^* < 0$. For $y \in [0,1]$, $G(y) > G(y^*) = 0$, $F'(z)|_{z=0} = G(y) > 0$, $F'(z)|_{z=1} = -G(y) < 0$, meaning that $z = 1$ is the stable strategy.

Case 2: When $C_E + \beta L_E - R_E > 0$, $p^* > 0$. When $p > p^*$, $x^* < 0$. For $x \in [0,1]$, the probability of $x > x^*$ is 1. Therefore, $y^* < 0$. For $y \in [0,1]$, $G(y) > G(y^*) = 0$, $F'(z)|_{z=0} = G(y) > 0$, $F'(z)|_{z=1} = -G(y) < 0$, thus $z = 1$ is the stable strategy. When $p < p^*$, $x^* > 0$. If $x > x^*$, $y^* < 0$. For $y \in [0,1]$, $G(y) > G(y^*) = 0$, $F'(z)|_{z=0} = G(y) > 0$, $F'(z)|_{z=1} = -G(y) < 0$, thus $z = 1$ is the stable strategy. If $x < x^*$, $y^* > 0$. When $y > y^*$, $G(y) > G(y^*) = 0$, $F'(z)|_{z=0} = G(y) > 0$, $F'(z)|_{z=1} = -G(y) < 0$, thus $z = 1$ is the stable strategy. When $y < y^*$, $G(y) < G(y^*) = 0$, $F'(z)|_{z=0} = G(y) < 0$, $F'(z)|_{z=1} = -G(y) > 0$, thus $z = 0$ is the stable strategy. When $y = y^*$, $G(y) = G(y^*) = 0$, $F'(z)|_{z=0} = F'(z)|_{z=1} = 0$, enterprises cannot determine a stable strategy. □

Proposition 3 suggests that the strategic choices of enterprises are influenced by government departments, digital platforms, and elderly people. When elderly people refuse to participate in data sharing, companies are unable to obtain effective demand-side information from digital platforms, thus reducing their willingness to cooperate. In this situation, support from government departments is a strong driving force for enterprises to choose a cooperation strategy. With the increase in government departments support, the willingness of enterprises to cooperate also increases. With improved regulatory measures, digital platforms tend to provide high-quality protection for enterprise data, enabling enterprises to obtain sufficient benefits from cooperation even without government support. When enterprises have high motivation to cooperate, government departments can gradually reduce subsidies and let the market develop independently.

The phase diagram of enterprises is shown in Figure 6.

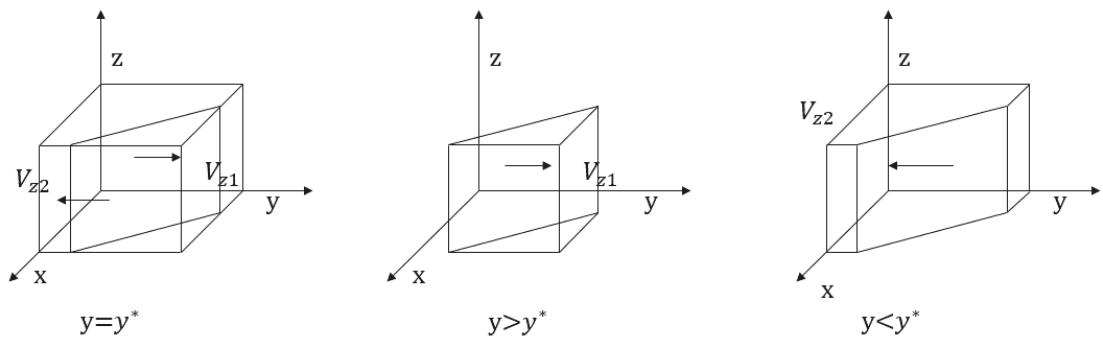


Figure 6. The phase diagram of enterprises.

Based on the above analysis, we solve for V_{z1} and V_{z2}.

$$V_{z2} = \int_0^1 \int_0^1 y^* dxdz = \frac{2C_E + 2\beta L_E - 2R_E - B_E - 2p\Delta R_E}{2(\beta - \alpha)L_E} \tag{20}$$

$$V_{z1} = 1 - V_{z2} = \frac{2R_E + B_E + 2p\Delta R_E - 2C_E - 2\alpha L_E}{2(\beta - \alpha)L_E} \tag{21}$$

Corollary 3. *The probability of enterprises choosing the cooperation strategy is positively correlated with cooperation benefits, government subsidies, and additional economic benefits brought by elderly people, and negatively correlated with cooperation costs, the probability of data leakage, and data leakage losses.*

Proof. V_{z1} and V_{z2} represent the probabilities of enterprises choosing cooperation strategy and non-cooperation strategy, respectively. Calculate the partial derivative V_{z1} with respect to $R_E, B_E, \Delta R_E, C_E, \beta$: $\frac{\partial V_{z1}}{\partial R_E} = \frac{1}{(\beta-\alpha)L_E} > 0$, $\frac{\partial V_{z1}}{\partial B_E} = \frac{1}{2(\beta-\alpha)L_E} > 0$, $\frac{\partial V_{z1}}{\partial \Delta R_E} = \frac{p}{(\beta-\alpha)L_E} > 0$, $\frac{\partial V_{z1}}{\partial C_E} = \frac{-1}{(\beta-\alpha)L_E} < 0$, $\frac{\partial V_{z1}}{\partial \beta} = -\frac{2R_E+B_E+2p\Delta R_E-2C_E-2\alpha L_E}{2(\beta-\alpha)^2 L_E} < 0$. □

The economic benefits of data sharing, government subsidies, and additional benefits brought by elderly people all have a positive impact on enhancing the willingness of enterprises to cooperate with digital platforms. High cooperation cost and potential data breach losses are important reasons that prevent enterprises from participating in silver economy data sharing. Therefore, choosing reputable digital platforms for collaboration will significantly reduce the likelihood of data breaches and thus lower costs. Moreover, large-scale digital platforms also have more customers, enabling enterprises to obtain more effective demand-side information to adjust their resource allocation and provide better services. In addition, government departments should also alleviate the economic burden on enterprises participating in data sharing and increase their enthusiasm by providing subsidies, tax reductions, and other incentives.

3.3.4. Stability Analysis of Strategies for Elderly People

The expected revenue for elderly people choosing P_1 (P_2) strategy is $M_{P1}(M_{P2})$, and the average expected revenue is M_P.

$$M_{P1} = R_P + xB_P - C_P - [\alpha y + \beta(1-y)]L_P \tag{22}$$

$$M_{P2} = 0 \tag{23}$$

$$M_P = p\{R_P + xB_P - C_P - [\alpha y + \beta(1-y)]L_P\} \tag{24}$$

The replicator dynamic equation and its first-order derivative are as follows:

$$F(p) = p(M_{P1} - M_P) = p(1-p)\{R_P + xB_P - C_P - [\alpha y + \beta(1-y)]L_P\} \tag{25}$$

$$F'(p) = \frac{d(F(p))}{dp} = (1-2p)[R_P + xB_P - C_P - [\alpha y + \beta(1-y)]L_P] \tag{26}$$

The probability of elderly people's decision-making in a stable state must satisfy $F(p) = 0$ and $F'(p) < 0$.

Proposition 4. *When $x > x^*$, elderly people tend to share data with digital platforms and enterprises. When $x < x^*$, if $y > y^*$, elderly people tend to share data with digital platforms and*

enterprises; if $y < y^*$, elderly people tend to not share data with digital platforms and enterprises. Here, y^* and x^* are as follows:

$$y^* = \frac{C_P + \beta L_P - R_P - xB_P}{(\beta - \alpha)L_P},\ x^* = \frac{C_P + \beta L_P - R_P}{B_P}$$

Proof. Let $G(y) = R_P + xB_P - C_P - [\alpha y + \beta(1-y)]L_P$. Since $\frac{\partial G(y)}{\partial y} = -(\alpha - \beta)L_P > 0$, $G(y)$ is an increasing function of y. Next, we discuss this proposition in two situations.

Case 1: When $C_P + \beta L_P - R_P < 0$, $x^* < 0$. When $x \in [0,1]$, the probability of $x > x^*$ is 1, then $y^* < 0$. Therefore, when $y \in [0,1]$,$G(y) > G(y^*) = 0$, $F'(p)|_{p=0} = G(y) > 0$, $F'(p)|_{p=1} = -G(y) < 0$, meaning that $p = 1$ is the stable strategy.

Case 2: When $C_P + \beta L_P - R_P > 0$, $x^* > 0$. If $x > x^*$, $y^* < 0$. Therefore, when $y \in [0,1]$, $G(y) > G(y^*) = 0$, $F'(p)|_{p=0} = G(y) > 0$, $F'(p)|_{p=1} = -G(y) < 0$, meaning that $p = 1$ is the stable strategy. If $x < x^*$, $y^* > 0$. When $y > y^*$,$G(y) > G(y^*) = 0$, $F'(p)|_{p=0} = G(y) > 0$, $F'(p)|_{p=1} = -G(y) < 0$, thus $p = 1$ is the stable strategy. When $y < y^*$, $G(y) < G(y^*) = 0$, $F'(p)|_{p=0} = G(y) < 0$, $F'(p)|_{p=1} = -G(y) > 0$, thus $p = 0$ is the stable strategy. When $y = y^*$, then $G(y) = 0$, $F'(p)|_{p=0} = G(y) = 0$, $F'(p)|_{p=1} = -G(y) = 0$, meaning that elderly people cannot determine a stable strategy. □

Proposition 4 indicates that the strategic choices of elderly people are influenced by government departments and digital platforms. Digital platforms' low willingness to provide high-quality protection exposes elderly people participating in data sharing to higher risks of privacy breaches. In such cases, government subsidies can encourage elderly people to shift from non-sharing to sharing data strategies. When digital platforms actively ensure data security, even if the government does not provide subsidies, elderly people will choose the sharing data strategy. Therefore, in the early stages of silver economy data sharing, government departments need to formulate reasonable subsidy policies to send positive signals to elderly people to encourage participation in data sharing. Meanwhile, digital platforms should actively express their demand and willingness to provide high-quality protection to encourage elderly people to adopt a data sharing strategy.

The phase diagram of elderly people is shown in Figure 7.

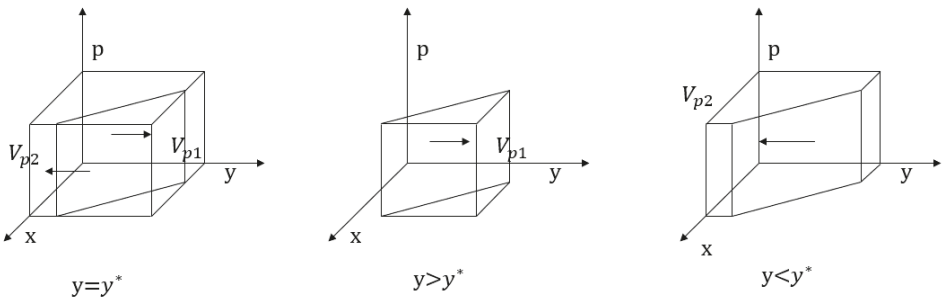


Figure 7. The phase diagram of elderly people.

Based on the above analysis, we solve for V_{p1} and V_{p2}.

$$V_{\mathrm{p2}} = \int_0^1 \int_0^1 y^* dx dp = \frac{2C_P + 2\beta L_P - 2R_P - xB_P}{2(\beta - \alpha)L_P} \tag{27}$$

$$V_{\mathrm{p1}} = 1 - V_{\mathrm{p2}} = \frac{xB_P + 2R_P - 2\alpha L_P - 2C_P}{2(\beta - \alpha)L_P} \tag{28}$$

Corollary 4. *The probability of elderly people choosing sharing data strategy is positively correlated with data sharing benefit and government subsidies, and negatively correlated with participation costs, the probability of data leakage, and data leakage losses.*

Proof. V_{p1} and V_{p2} represent the probability of elderly people choosing the sharing data strategy and the non-sharing data strategy, respectively. Calculate the partial derivative V_{p1} with respect to R_P, B_P, C_P, β, L_P: $\frac{\partial V_{p1}}{R_P} = \frac{1}{(\beta-\alpha L_P)} > 0$, $\frac{\partial V_{p1}}{B_P} = \frac{x}{(\beta-\alpha L_P)} > 0$, $\frac{\partial V_{p1}}{C_P} = \frac{-1}{(\beta-\alpha L_P)} < 0$, $\frac{\partial V_{p1}}{\beta} = \frac{-L_P}{2((\beta-\alpha)L_P)^2} < 0$, $\frac{\partial V_{p1}}{L_P} = \frac{-(xB_P+2R_P-2C_P)}{{L_P}^2} < 0$. □

As the benefits that elderly people can obtain from data sharing increase, government subsidies increase, and participation costs decreases, elderly people will be more inclined to actively participate in data sharing. Therefore, government departments can encourage elderly people to participate in data sharing by providing subsidies while improving their digital literacy through lectures and other means, thereby reducing the difficulty and cost of sharing data with other entities through digital devices. At the same time, government departments should raise elderly people's awareness of data protection to reduce losses caused by data leakage accidents.

4. Model Analysis

In the game system involving government departments, digital platforms, enterprises, and elderly people, the stability of strategy combination can be judged according to the Lyapunov's first law. In a dynamic game process, the equilibrium of a mixed strategy cannot reach a stable equilibrium state [50,51]. Therefore, this paper only needs to analyze the 16 pure equilibrium points [52]. The Jacobian matrix is as follows:

$$J = \begin{bmatrix} \frac{\partial F(x)}{\partial x} & \frac{\partial F(x)}{\partial y} & \frac{\partial F(x)}{\partial z} & \frac{\partial F(x)}{\partial p} \\ \frac{\partial F(y)}{\partial x} & \frac{\partial F(y)}{\partial y} & \frac{\partial F(y)}{\partial z} & \frac{\partial F(y)}{\partial p} \\ \frac{\partial F(z)}{\partial x} & \frac{\partial F(z)}{\partial y} & \frac{\partial F(z)}{\partial z} & \frac{\partial F(z)}{\partial p} \\ \frac{\partial F(p)}{\partial x} & \frac{\partial F(p)}{\partial y} & \frac{\partial F(p)}{\partial z} & \frac{\partial F(p)}{\partial p} \end{bmatrix} \tag{29}$$

4.1. Stability Analysis Under Passive Promotion of Government Departments

When government departments choose the passive promotion strategy, i.e., $-C_{Gh} + C_{Gl} + (\varphi - \omega)(1-y)\Delta P + (1-y)p\theta R_G < 0$, the stability analysis is shown in Table 2.

The strategic choices of digital platforms, enterprises, and elderly people are determined by their benefits and costs. By calculating the eigenvalues, when government departments choose the passive promotion strategy, all eight equilibrium points can become stable points. The analysis of the eight equilibrium points is as follows:

At E_1, E_3, E_4, and E_7, digital platforms tend to choose to provide low-quality protection due to the additional cost of choosing the high-quality protection strategy, which increases the risks of elderly people's and enterprises' data leakage and is not conductive to data sharing in the silver economy. Hence, measures should be taken to prevent these points from becoming ESS, such as strengthening government punishment, improving supervision efficiency by elderly people, and encouraging all parties to actively report security vulnerabilities in the data sharing process.

At the three equilibrium points E_2, E_5, and E_6, although digital platforms actively protect the security of elderly people's and enterprises' data, enterprises or elderly people refuse to participate in data sharing due to the inability to obtain sufficient profits. In the early stages, enterprises and elderly people need to invest huge costs and may not be able to obtain satisfactory benefits from data sharing. As the industry develops, the benefits

of participating in data sharing will increase. Therefore, government departments should address the issue of insufficient returns for enterprises and elderly people in the early stages by providing subsidies, tax reductions, and other means.

Table 2. Stability analysis under passive promotion of government departments.

Equilibrium Point	Eigenvalue λ_1,λ_2,λ_3,λ_4	Stability Judgment
$E_1(0,0,0,0)$	$-C_{Gh}+C_{Gl}+(\varphi-\omega)\Delta P, -C_{Dh}+C_{Dl}+\omega\Delta P-\alpha[(1-\lambda)P+L_D]+\beta[(1+\lambda)P+L_D], R_E-C_E-\beta L_E, R_P-C_P-\beta L_P$	ESS when $\lambda_2<0,\ \lambda_3<0$, and $\lambda_4<0$ are satisfied
$E_2(0,1,0,0)$	$-C_{Gh}+C_{Gl}, C_{Dh}-C_{Dl}-\omega\Delta P+\alpha[(1-\lambda)P+L_D]-\beta[(1+\lambda)P+L_D], R_E-C_E-\alpha L_E, R_P-C_P-\alpha L_P$	ESS when $\lambda_2<0,\ \lambda_3<0$, and $\lambda_4<0$ are satisfied
$E_3(0,0,1,0)$	$-C_{Gh}+C_{Gl}+(\varphi-\omega)\Delta P+p\theta R_G, -C_{Dh}+C_{Dl}+\omega\Delta P-\alpha[(1-\lambda)P+L_D]+\beta[(1+\lambda)P+L_D], -R_E+C_E+\beta L_E, R_P-C_P-\beta L_P$	ESS when $\lambda_2<0,\ \lambda_3<0$, and $\lambda_4<0$ are satisfied
$E_4(0,0,0,1)$	$-C_{Gh}+C_{Gl}+(\varphi-\omega)\Delta P, -C_{Dh}+C_{Dl}+\omega\Delta P-\alpha[(1-\lambda)P+L_D]+\beta[(1+\lambda)P+L_D]+\theta S, R_E+\Delta R_E-C_E-\beta L_E, -R_P+C_P+\beta L_P$	ESS when $\lambda_2<0,\ \lambda_3<0$, and $\lambda_4<0$ are satisfied
$E_5(0,1,1,0)$	$-C_{Gh}+C_{Gl}, C_{Dh}-C_{Dl}-\omega\Delta P+\alpha[(1-\lambda)P+L_D]-\beta[(1+\lambda)P+L_D], -R_E+C_E+\alpha L_E, R_P-C_P-\alpha L_P$	ESS when $\lambda_2<0,\ \lambda_3<0$, and $\lambda_4<0$ are satisfied
$E_6(0,1,0,1)$	$-C_{Gh}+C_{Gl}, C_{Dh}-C_{Dl}-\omega\Delta P+\alpha[(1-\lambda)P+L_D]-\beta[(1+\lambda)P+L_D]-\theta S, R_E+\Delta R_E-C_E-\alpha L_E, -R_P+C_P+\alpha L_P$	ESS when $\lambda_2<0,\ \lambda_3<0$, and $\lambda_4<0$ are satisfied
$E_7(0,0,1,1)$	$-C_{Gh}+C_{Gl}+(\varphi-\omega)\Delta P+p\theta R_G, -C_{Dh}+C_{Dl}+\omega\Delta P-\alpha[(1-\lambda)P+L_D]+\beta[(1+\lambda)P+L_D]+\theta S, -R_E-\Delta R_E+C_E+\beta L_E, -R_P+C_P+\beta L_P$	ESS when $\lambda_2<0,\ \lambda_3<0$, and $\lambda_4<0$ are satisfied
$E_8(0,1,1,1)$	$-C_{Gh}+C_{Gl}, C_{Dh}-C_{Dl}-\omega\Delta P+\alpha[(1-\lambda)P+L_D]-\beta[(1+\lambda)P+L_D]-\theta S, -R_E-\Delta R_E+C_E+\alpha L_E, -R_P+C_P+\alpha L_P$	ESS when $\lambda_2<0,\ \lambda_3<0$, and $\lambda_4<0$ are satisfied

At point E_8, digital platforms choose to provide high-quality protection, enterprises choose to cooperate with platforms, and elderly people choose to share data with platforms and enterprises. At this point, the system reaches a Pareto optimal state, which is also the ideal strategy combination of the system. When the expected revenue from supplying high-quality protection is greater than that from low-quality protection, and the benefits obtained by enterprises and elderly people from silver economy data sharing are greater than the sum of costs and potential losses, E_8 can become a stable point in the system.

4.2. Stability Analysis Under Positive Promotion of Government Departments

When government departments choose the positive promotion strategy, i.e., $-C_{Gh}+C_{Gl}+(\varphi-\omega)(1-y)\Delta P+(1-y)p\theta R_G>0$, the stability analysis is shown in Table 3.

As shown in Table 3, there are no pure-strategy stable equilibrium points where government departments choose the positive promotion strategy. When government departments choose the positive promotion strategy and digital platforms tend to provide high-quality protection, the prerequisite is $-C_{Gh}+C_{Gl}>0$, which contradicts the model assumption that $C_{Gh}>C_{Gl}$. If government departments maintain a stable positive promotion strategy with well-implemented regulatory and punitive measures for digital platforms, digital platforms will choose to provide high-quality protection. However, when digital platforms choose the high-quality protection strategy, the government departments' positive promotion strategy becomes unstable, which is consistent with the previous stability analysis results. As digital platforms increase their investments and actively ensure data security, government departments may opt for the passive promotion strategy to reduce administrative costs.

Based on the above analysis, the following can be concluded:

1. To ensure the long-term safeguarding of data security in the silver economy data-sharing process and protect the interests of enterprises and elderly people, the government needs to bear additional regulatory costs.
2. When government departments enforce positive promotion, regulatory and punitive mechanisms for digital platforms incentivize platforms to provide high-quality protection of user data, facilitating a virtuous cycle of development in silver economy data sharing.

Table 3. Stability analysis under positive promotion of government departments.

Equilibrium Point	Eigenvalue λ_1,λ_2,λ_3,λ_4	Stability Judgment
$E_9(1,0,0,0)$	$C_{Gh}-C_{Gl}-(\varphi-\omega)\Delta P, -C_{Dh}+C_{Dl}+\varphi\Delta P-\alpha[(1-\lambda)P+L_D]+\beta[(1+\lambda)P+L_D], R_E+B_E-C_E-\beta L_E, R_P+B_P-C_P-\beta L_P$	US
$E_{10}(1,1,0,0)$	$C_{Gh}-C_{Gl}, C_{Dh}-C_{Dl}-\varphi\Delta P+\alpha[(1-\lambda)P+L_D]-\beta[(1+\lambda)P+L_D], R_E+B_E-C_E-\alpha L_E, R_P+B_P-C_P-\alpha L_P$	$\otimes$
$E_{11}(1,0,1,0)$	$C_{Gh}-C_{Gl}-(\varphi-\omega)\Delta P, -C_{Dh}+C_{Dl}+\varphi\Delta P-\alpha[(1-\lambda)P+L_D]+\beta[(1+\lambda)P+L_D], -R_E-B_E+C_E+\beta L_E, R_E+B_E-C_E-\beta L_E$	US
$E_{12}(1,0,0,1)$	$C_{Gh}-C_{Gl}-(\varphi-\omega)\Delta P-R_G, -C_{Dh}+C_{Dl}+\varphi\Delta P-\alpha[(1-\lambda)P+L_D]+\beta[(1+\lambda)P+L_D]+\theta S, R_E+B_E+\Delta R_E-C_E-\beta L_E, -R_P-B_P+C_P+\beta L_P$	US
$E_{13}(1,1,1,0)$	$C_{Gh}-C_{Gl}, C_{Dh}-C_{Dl}-\varphi\Delta P+\alpha[(1-\lambda)P+L_D]-\beta[(1+\lambda)P+L_D], -R_E-B_E+C_E+\alpha L_E, R_P+B_P-C_P-\alpha L_P$	$\otimes$
$E_{14}(1,1,0,1)$	$C_{Gh}-C_{Gl}, C_{Dh}-C_{Dl}-\varphi\Delta P+\alpha[(1-\lambda)P+L_D]-\beta[(1+\lambda)P+L_D]-\theta S, R_E+B_E+\Delta R_E-C_E-\alpha L_E, -R_P-B_P+C_P+\alpha L_P$	$\otimes$
$E_{15}(1,0,1,1)$	$C_{Gh}-C_{Gl}-(\varphi-\omega)\Delta P-R_G, -C_{Dh}+C_{Dl}+\varphi\Delta P-\alpha[(1-\lambda)P+L_D]+\beta[(1+\lambda)P+L_D]+\theta S, -R_E-B_E-\Delta R_E+C_E+\beta L_E, -R_P-B_P+C_P+\beta L_P$	US
$E_{16}(1,1,1,1)$	$C_{Gh}-C_{Gl}, C_{Dh}-C_{Dl}-\varphi\Delta P+\alpha[(1-\lambda)P+L_D]-\beta[(1+\lambda)P+L_D]-\theta S, -R_E-B_E-\Delta R_E+C_E+\alpha L_E, -R_P-B_P+C_P+\alpha L_P$	$\otimes$

US, unstable point; $\otimes$, equilibrium point cannot fulfill the prerequisites.

5. Numerical Simulations and Analysis

The system composed of government departments, digital platforms, enterprises, and elderly people exhibits different evolutionary paths in silver economy data sharing at different stages. To verify the accuracy of the results of the theoretical analysis and analyze the influence of parameters, this paper uses MATLAB 2023a to conduct numerical simulations.

The model encompasses a multitude of parameters that are difficult to accurately obtain in practical applications, and different parameter values may lead to significant differences in results. Research has shown that the value of a simulation model lies in its explanatory power and ability to explain mechanisms, rather than in fully replicating reality [53]. According to the *Annual Report on the Development of Silver Economy in China (2024)* [54], the current scale of China's silver economy is around CNY 7 trillion, and annual government subsidies for elderly care services are about CNY 70 billion. In addition, existing evolutionary game studies on data sharing [4,36,45] provide useful guidance for calibrating the relative scales of costs, penalties, and benefits among various parties in the model. Parameter settings under different scenarios are as follows, ensuring that the parameter values do not deviate from industry benchmarks.

5.1. Numerical Simulation Analysis of Equilibrium Points

This section mainly verifies the accuracy of the theoretical analysis results. This article selects two extreme scenarios within the system, Scenario 1 and Scenario 8, for numerical simulation analysis.

5.1.1. Numerical Simulation of Case 1

To meet the condition that $-C_{Dh} + C_{Dl} + \omega\Delta P - \alpha[(1-\lambda)P + L_D] + \beta[(1+\lambda)P + L_D] < 0, R_E - C_E - \beta L_E < 0, R_P - C_P - \beta L_P < 0$, the parameters of the model are set as follows: $C_{Gh} = 55$, $C_{Gl} = 50$, $C_{Dh} = 35$, $C_{Dl} = 15$, $\varphi = 0.8$, $\omega = 0.3$, $\alpha = 0.2$, $\beta = 0.5$, $\lambda = 0.4$, $\Delta P = 15$, $P = 10$, $R_E = 30$, $C_E = 25$, $R_P = 5$, $C_P = 4$. $L_D = 25$, $L_E = 20$, $L_P = 3$, $\Delta R_E = 10$, $\theta = 0.3$, $S = 5$.

The final evolution result is shown in Figure 8. The system eventually evolves to the equilibrium point $E_1(0,\ 0,\ 0,\ 0)$; that is, government departments choose passive promotion, digital platforms choose low-quality protection, enterprises choose non-cooperation, and elderly people choose non-sharing. At this time, elderly people are not aware of the benefits that participating in data sharing can bring and lack active guidance from the government; therefore, their willingness to do so is relatively low. For enterprises, the benefits of cooperating with digital platforms are less than operating and management expenses, so they are more inclined not to cooperate with them. In the absence of relevant laws and regulations and strict government supervision, digital platforms provide low-quality protection for user data to save costs. Data leakage accidents often occur, which further hinders the development of data sharing in the silver economy.

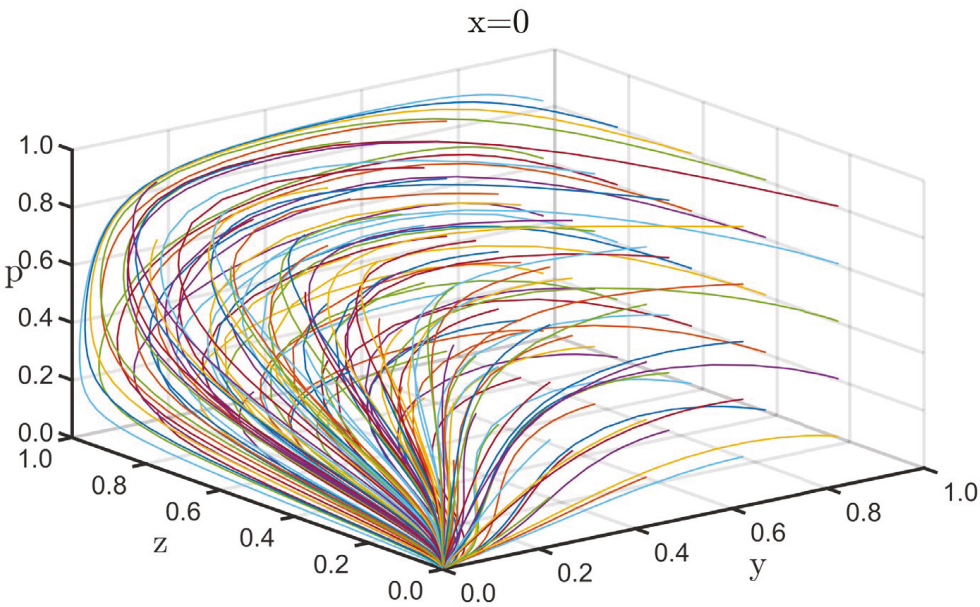


Figure 8. Simulation result of Case 1.

To address these challenges, it is imperative for the government to increase investments in subsidies and tax reductions to incentivize participation from both enterprises and elderly users, thereby lowering their data-sharing costs. Simultaneously, comprehensive legal regulations and strict supervision over digital platforms' practices will safeguard user data security and ensure the progression of silver economy data sharing.

5.1.2. Numerical Simulation of Case 8

To satisfy the inequality conditions of Case 8, we adjusted $C_{Dh} = 30$ and kept all other parameters consistent with Case 1.

The results obtained are shown in Figure 9. The system eventually evolves to the equilibrium point $E_8(0,\ 1,\ 1,\ 1)$; that is, government departments choose passive promotion, digital platforms choose low-quality protection, enterprises choose cooperation, and elderly people choose sharing data.

At this point, the system reaches a Pareto optimal state, and enterprises and elderly people can obtain stable benefits from participating in data sharing. In this state, there is no need for active government intervention, and data sharing of the silver economy can proceed smoothly. For digital platforms, the cost savings from opting to provide low-quality protection strategies are outweighed by the direct economic losses and reputational damages caused by potential data breaches. As a result, digital platforms receive incentives to implement proactive measures to protect user data.

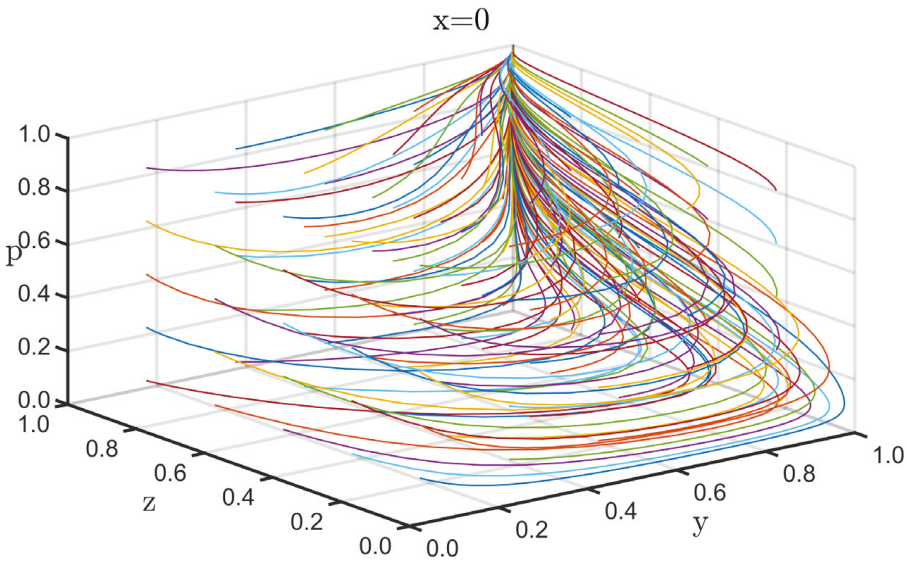


Figure 9. Simulation result of Case 8.

5.2. Simulation Analysis of the Changes in Initial Probability

To study the influence of the initial strategy probabilities of each party on evolutionary paths, the parameter settings are as follows: $C_{Gh} = 55, C_{Gl} = 50, C_{Dh} = 35, C_{Dl} = 15$, $\varphi = 0.8, \omega = 0.3, \alpha = 0.2, \beta = 0.5, \lambda = 0.4$, $\Delta P = 10$, $P = 10$, $R_E = 30$, $C_E = 25$, $R_P = 5$, $C_P = 4$, $L_D = 25$, $L_E = 20$, $L_P = 3$, $\Delta R_E = 10$, $\theta = 0.3$, $S = 10$. These parameter settings simultaneously satisfy both Case 1 and Case 8, meaning that during the analysis of the impact of initial strategy probabilities on evolutionary paths, the system should exhibit two stable equilibrium points. The initial strategies of government departments, digital platforms, enterprises, and elderly people are set as $(x, y, z, p) = (0.2, 0.2, 0.2, 0.2)$, $(0.5, 0.5, 0.5, 0.5)$, $(0.8, 0.8, 0.8, 0.8)$.

Using three initial strategy combinations for simulation, the dynamic evolution process of the system is shown in Figure 10. When the initial probabilities of proactive strategies among the four subjects are low, the system evolves to $E_1(0, 0, 0, 0)$. As the initial proactive strategy probabilities increase, the system transitions from $E_1(0, 0, 0, 0)$ to $E_8(0, 1, 1, 1)$. Under identical parameters, changes in participants' initial strategy proportions alter the system's evolutionary path. The higher the initial proportion of proactive strategies among all parties, the greater the probability of the system converging to the Pareto optimal state.

Therefore, effectively enhancing the proactiveness of all stakeholders in participating in data sharing is crucial for the development of the silver economy industry. Government departments can achieve this goal by intensifying publicity efforts and providing subsidies. These measures lower participation barriers and align individual interests with collective welfare, fostering a virtuous cycle of data-driven innovation and inclusive growth in aging societies.

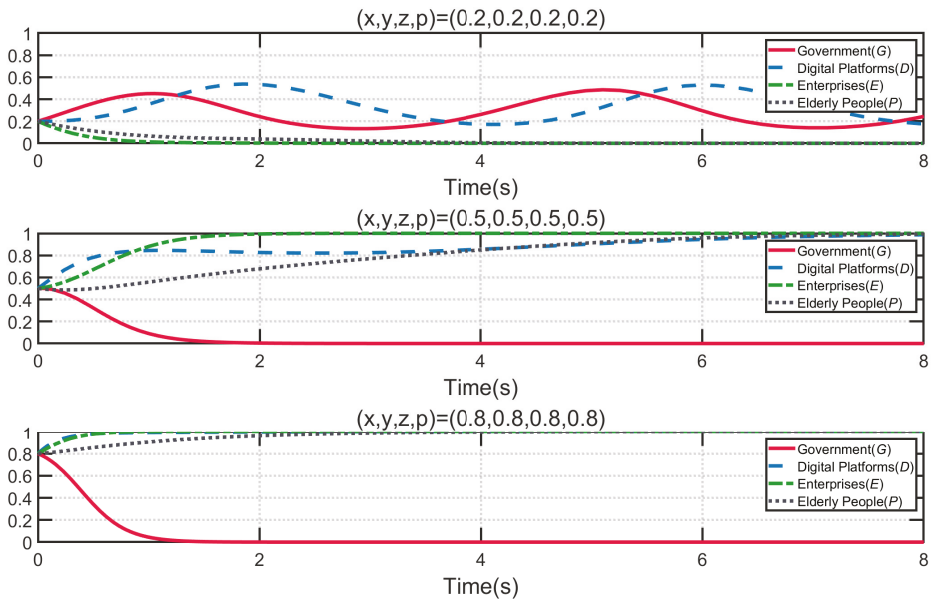


Figure 10. The impact of the initial strategy of the four parties.

5.3. Simulation Analysis of Parameters

The stability of each subject's strategy is susceptible to external parameters. In order to study how to achieve the ideal state of the system, where digital platforms provide high-quality protection and enterprises and users participate in data sharing, this section assigns parameters that can meet the conditions of Case 8. The initial strategy of government departments, digital platforms, enterprises, and elderly people are set as $(x, y, z, p) = (0.5, 0.3, 0.2, 0.3)$.

5.3.1. The Impact of Penalties (P) on Quadratic Evolutionary Behavior

Keeping other parameters unchanged and changing government penalties on digital platforms by taking $P = \{10, 15, 20\}$, the evolutionary path of each subject is obtained as in Figure 11.

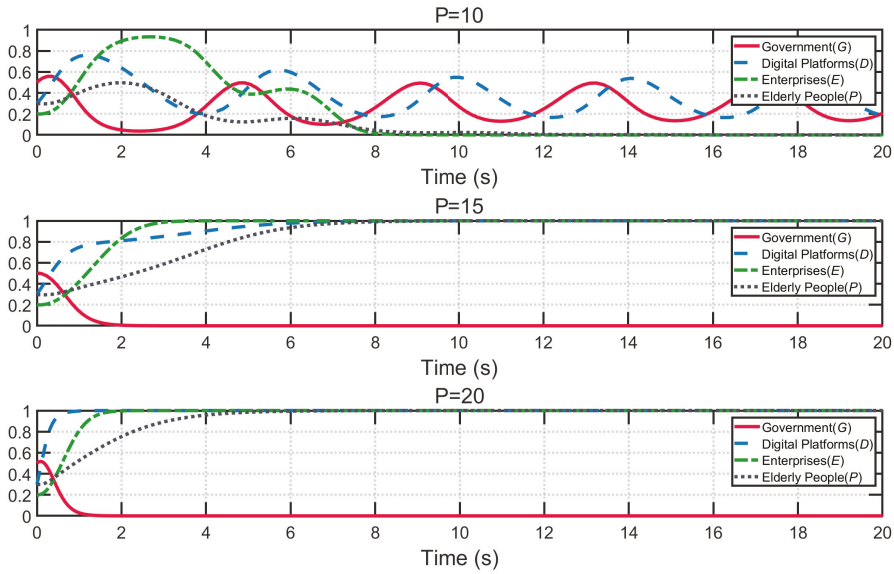


Figure 11. The impact of penalties (P) on quadratic evolutionary behavior.

As illustrated in the figure, as government departments increase penalties on digital platforms that adopt low-quality protection leading to data breaches, platform cost rises,

incentivizing them to provide high-quality protection. When P exceeds a critical threshold, the probability of platforms choosing high-quality protection reaches 1. Moreover, higher P accelerates the transition speed of platforms toward offering high-quality protection, which in turn drives faster participation by enterprises and elderly people in data sharing. Concurrently, to reduce costs, government departments gradually shift from positive promotion to passive promotion.

To ensure data security and promote data sharing in the silver economy, government departments should strengthen regulatory mechanisms and increase penalties for digital platforms that cause data leakage accidents.

At the same time, government departments also need to consider the risks associated with the transmission of highly sensitive data, as relying solely on digital platforms cannot fully defend against malicious network attacks. Once the government's punishment is too severe, digital platforms will reduce their enthusiasm to participate in data sharing due to the inability to afford high fines. Therefore, excessive government regulation and intervention can also affect normal market operation.

5.3.2. The Impact of Complaint Probability (θ) on Quadratic Evolutionary Behavior

Keeping the other parameters unchanged and changing the probability of elderly people filing complaints against digital platforms that provide low-quality protection, let $\theta = \{0.3, 0.5, 0.6\}$. The strategy evolution process of the four-party game participants is shown in Figure 12.

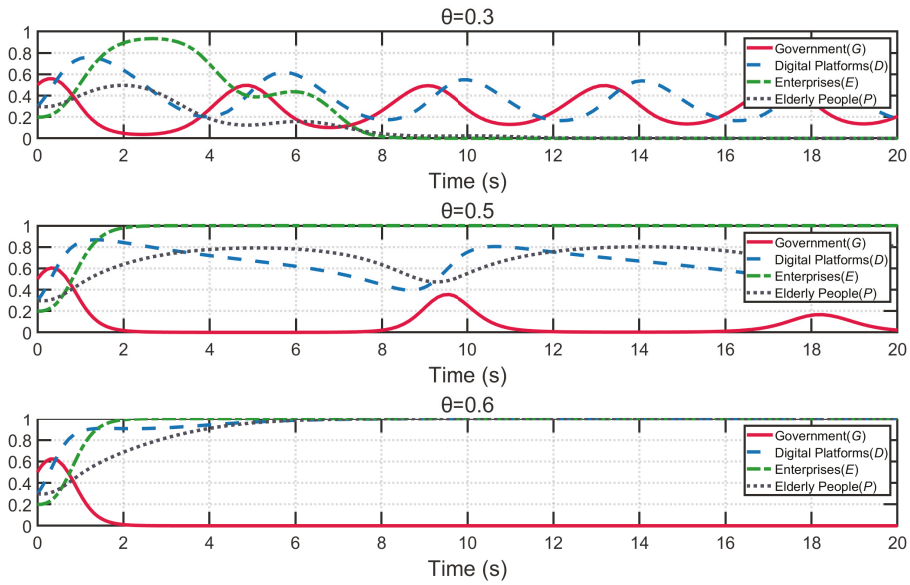


Figure 12. The impact of complaint probability (θ) on quadratic evolutionary behavior.

As illustrated in the figure, with the improvement of digital literacy, elderly people become more likely to detect oversights in digital platforms' data collection processes, leading to increased probabilities of lodging complaints. To avoid reputational losses caused by excessive complaints, digital platforms will standardize their operational practices, gradually raising the probability of providing high-quality protection, while enterprises and elderly people will be more inclined to participate in data sharing to obtain greater benefits.

In summary, the complaint probability θ of elderly people significantly impacts the decision-making of digital platforms. Enhancing the digital literacy of the elderly is key to helping them identify oversights in digital platforms' operations. It provides the government with new insights for promoting silver economy data sharing: improving the digital literacy of elderly populations through science popularization lectures and

diversified digital skills training. These measures will enable elderly people to detect issues during data sharing, provide timely feedback, safeguard their interests, and drive the development of the silver economy industry.

5.3.3. The Impact of Reputational Loss (S) on Quadratic Evolutionary Behavior

Keeping other parameters unchanged and changing the reputational loss S incurred by digital platforms due to complaints from elderly people, let $S = \{5, 8, 10\}$. The evolutionary path of each subject is shown in Figure 13.

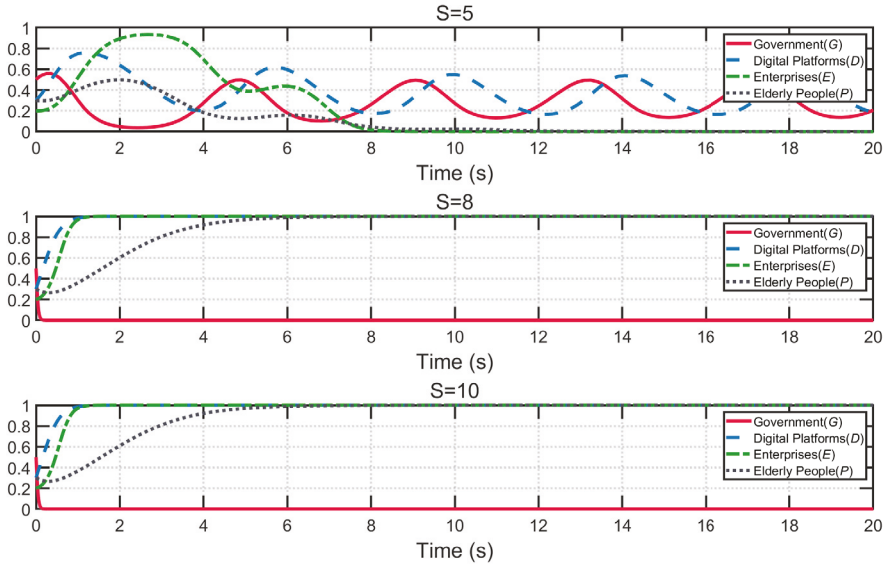


Figure 13. The impact of reputational loss (S) on quadratic evolutionary behavior.

As illustrated in the figure, as the reputational loss caused by complaints from enterprises and elderly people against digital platforms increases, the probability of digital platforms choosing to provide high-quality protection gradually rises, so enterprises and elderly people are more likely to participate in silver economy data sharing.

Hence, the magnitude of reputational impact resulting from elderly people's complaints influences the decisions of digital platforms. Professional and reputable digital platforms place greater emphasis on the user experience of elderly individuals, as complaints would impose larger reputational losses. Consequently, such digital platforms are more likely to adopt active security measures for user data. Therefore, for government departments, enterprises, and elderly people, collaborating with more established and specialized digital platforms is more advantageous for safeguarding their interests. For digital platforms, prioritizing user experience and maintaining a positive public image and reputation will attract more users to collaborate, thereby securing greater benefits.

5.3.4. The Impact of Penalty Coefficient (λ) on Quadratic Evolutionary Behavior

Keeping other parameters unchanged and changing the penalty coefficient λ imposed by the government departments on digital platforms, let $\lambda = \{0.4, 0.5, 0.6\}$. The strategic evolution of the four-party game participants is depicted in Figure 14.

The figure demonstrates that an increase in λ amplifies the penalty differential between fines levied on digital platforms for data breaches under providing high-quality protection and low-quality protection. If the cost savings from adopting low-quality protection are exceeded by the heightened penalty gap resulting from data breaches, platforms exhibit a stronger preference for offering high-quality protection. This behavioral shift drives enterprises to cooperate and encourages elderly people to share data with digital platforms.

Consequently, calibrating an optimal penalty adjustment parameter λ serves as a critical lever for advancing data sharing in the silver economy.

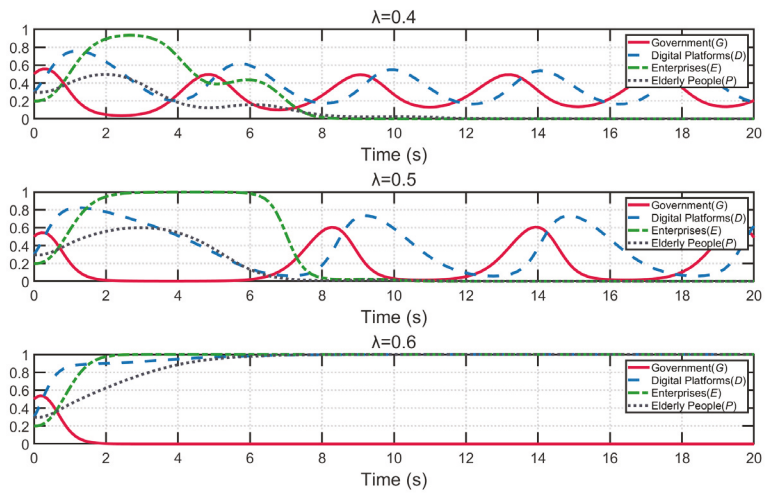


Figure 14. The impact of penalty coefficient (λ) on quadratic evolutionary behavior.

5.3.5. The Impact of Enterprise Participation Cost (C_E) on Data Sharing

Keeping other parameters unchanged and changing the enterprise participation cost C_E, let $C_E = \{20, 25, 30, 35\}$. The strategy evolution processes of digital platforms, enterprises, and elderly people are shown in Figure 15.

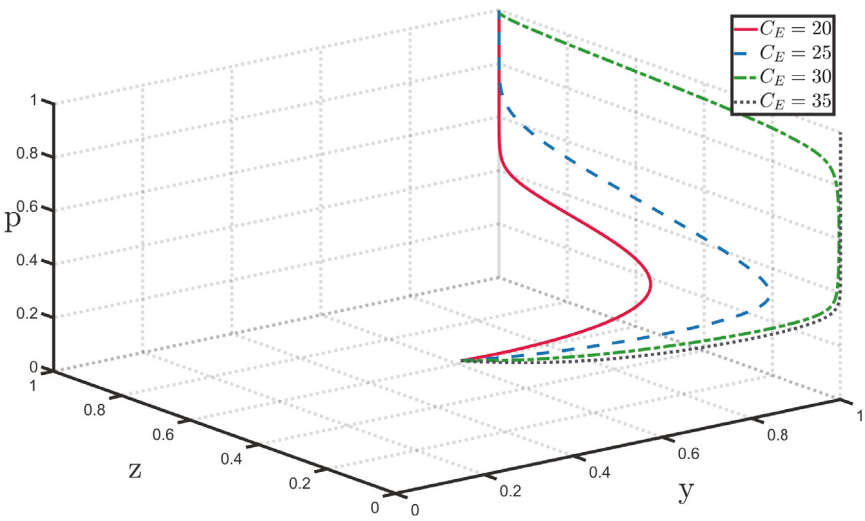


Figure 15. The impact of enterprise participation cost (C_E) on data sharing.

As the cost of enterprise participation in silver economy data sharing increases, the speed at which the system evolves toward the equilibrium point $E_8(0, 1, 1, 1)$ gradually slows. When the cost C_E exceeds a certain threshold, the stable point shifts from $E_8(0, 1, 1, 1)$ to $E_6(0, 1, 0, 1)$. This indicates that as C_E rises, enterprises' willingness to participate in data sharing progressively declines.

The composition of data-sharing costs C_E varies across enterprises of different scales. For small and medium enterprises (SMEs), the costs primarily arise from operational expenses and the establishment and maintenance of data-sharing channels. Compared to SMEs, large enterprises possess greater data exclusivity advantages, as the attenuation of competitive advantages due to data sharing is more pronounced [55]. Consequently, their willingness to engage in silver economy data sharing is lower than that of SMEs.

Governments can tailor incentive policies by enterprise scale, by prioritizing technical support for SMEs to lower barriers in establishing and maintaining data-sharing channels, thereby reducing their participation costs. On the other hand, for large enterprises, their competitive advantage may weaken due to sharing high-value and exclusive datasets. Therefore, strategic compensation through tax exemptions or subsidies is crucial for offsetting implicit losses.

5.3.6. The Impact of Elderly People's Participation Cost (C_P) on Data Sharing

Keeping other parameters unchanged and changing the elderly people's participation cost C_P, let $C_P = \{10, 11, 12, 13\}$. The strategy evolution processes of digital platforms, enterprises, and elderly people are shown in Figure 16.

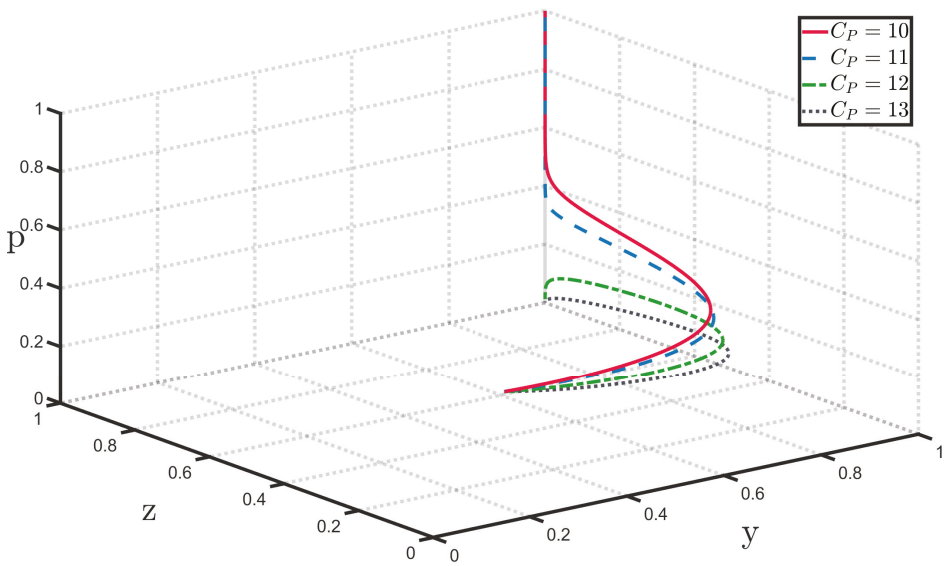


Figure 16. The impact of elderly people's participation cost (C_P) on data sharing.

As the cost C_P for elderly people participating in silver economy data sharing increases, the system's convergence speed toward $E_8(0, 1, 1, 1)$ gradually slows. When C_P exceeds a certain threshold, the stable equilibrium shifts from $E_8(0, 1, 1, 1)$ to $E_5(0, 1, 1, 0)$, indicating that higher C_P progressively reduces elderly people's willingness to engage in data sharing.

The costs for elderly people primarily include operational barriers and expenses related to purchasing smart devices. To address these challenges, government departments can implement fiscal subsidies, conduct digital literacy workshops, and organize community training programs, while digital platforms should develop elderly-adaptive application interfaces and strengthen data protection measures to jointly create a secure and trustworthy participation environment, thereby enhancing elderly people's willingness to share data.

6. Discussion and Conclusions

6.1. Discussion

This paper focuses on the context of population aging and explores how digital platforms can be utilized as intermediaries under government guidance to promote national-level silver economy data sharing. We establish a four-party game model to analyze silver economy data sharing, involving government departments, digital platforms, enterprises, and elderly people. Equilibrium points are found using differential equations, their stability is checked, and numerical simulations with MATLAB are performed to see how initial

strategies and external factors affect stakeholders' decisions. The main conclusions are as follows.

1. The four-party equilibrium game model reveals eight asymptotically stable points. When specific conditions are met—namely, when economic losses, fines, and reputational damages from data breaches under digital platforms' low-quality protection exceed their cost savings; when enterprises' benefits from cooperation outweigh cooperation costs and potential data breach losses; and when elderly people's benefits from data sharing surpass participation costs and potential risks—the system evolves toward (passive promotion, high-quality protection, cooperation, sharing) as the evolutionarily stable strategy. This state constitutes a Pareto optimal scenario where all parties voluntarily engage in data sharing without government intervention, although achieving this equilibrium in practice remains challenging.
2. Numerical simulations demonstrate that when parameters align with Case 1 and Case 8 conditions, the system concurrently sustains two stable strategies. The convergence outcome depends on the initial strategy proportions of the four parties. When stakeholders have a higher initial percentage of proactive strategies, they are more likely to reach the Pareto optimal state, whereas lower proportions increase the likelihood of convergence to the Pareto worst-case state.
3. Digital platforms, as the core actors in silver economy data sharing, critically influence the behavioral decisions of the other three parties. When platforms provide high-quality protection, government departments can reduce costs through passive promotion, and enterprises and elderly people actively engage in data sharing. Additionally, platforms' strategies are reciprocally shaped by the behaviors of government departments, enterprises, and elderly people. Higher fines and penalty adjustment parameters for data breaches under low-quality protection incentivize platforms to prioritize active data security. Elderly people with advanced digital literacy are more likely to identify platform vulnerabilities, leading to higher complaint probabilities, which impose significant reputational losses and compel platforms to adopt high-quality protection. Established, professional platforms prioritize user experience, making them more sensitive to reputational damages from complaints compared to smaller or less reputable platforms.

6.2. Conclusions

Based on the research results, we propose some policy recommendations to promote data sharing and the development of the silver economy. The suggestions are as follows:

1. Currently, China's silver economy remains in its nascent stage. As the guide and facilitator of silver economy data sharing, the government should encourage greater participation from enterprises and elderly people by fostering a robust data-sharing ecosystem. Simulation experiments indicate that the initial proportion of participants adopting proactive strategies—i.e., their baseline willingness to engage—is a critical determinant of whether the system achieves a Pareto optimal state. Government departments should strengthen top-level design for the silver economy data market, promptly enact relevant laws and regulations, expand public awareness campaigns to alleviate stakeholder concerns, and cultivate a trustworthy environment. Concurrently, offering incentives such as tax reductions and subsidies to participating enterprises and elderly people can enhance their interest. These measures will improve the efficiency of data sharing and amplify its societal benefits.
2. Government departments should rationally calibrate regulatory intensity and penalties for digital platforms. Establishing a comprehensive and adaptive regulatory framework can identify operational deficiencies in digital platforms, reduce the likeli-

hood of data breaches during sharing, and safeguard stakeholder interests. Government departments should simultaneously set up different punishment mechanisms, such as imposing high fines on platforms that adopt low-quality protection strategies leading to data breaches, and introducing penalty adjustment parameters to incentivize proactive behavior while curbing negligent behavior throughout the industry.

3. The cost of enterprise participation in data sharing is a pivotal factor influencing engagement. Reducing these costs effectively increases participation likelihood. Government departments should implement differentiated incentives based on enterprise scale: for SMEs, provide technical support to lower barriers in establishing data-sharing infrastructure; for large enterprises, offer tax relief to offset implicit competitive losses incurred from data sharing.
4. To advance silver economy data sharing, government departments should elevate elderly users' digital literacy through educational workshops and diversified skills training. Enhanced digital literacy empowers elderly people to protect their privacy and assert their interests during data sharing. Additionally, complaints and feedback from elderly people regarding platform security flaws impose reputational costs on platforms, compelling them to prioritize high-quality data protection.
5. The professionalism and credibility of digital platforms are key determinants of collaboration decisions by enterprises and elderly people. Government departments can periodically recognize and endorse exemplary platforms, incentivizing them to enhance data security and user experience to earn official recommendations. This approach fosters a virtuous competitive environment where platforms compete to meet higher standards, driving industry-wide improvement.

At the same time, government departments also need to pay attention to the ethical issues brought about by silver economy data sharing. Excessive focus on economic development and relaxed regulation will condone digital platforms and enterprises to excessively commercialize data, forming data monopolies, and infringing on the privacy rights of the elderly. Conversely, excessive intervention in market mechanisms may undermine market fairness and trigger controversies over the excessive capture of private data by public power. Therefore, data privacy protection is the result of joint participation from multiple parties. In addition to government legislation and regulation, enterprises also need to take responsibility and ensure compliant operations; digital platforms should actively respond to modern threats such as cyberattacks and prevent data breaches; and elderly people should raise their awareness of privacy protection and actively safeguard their data rights.

While this study offers valuable insights, additional research is necessary. Firstly, this article assumes that all subjects' strategy choices are binary and static; future research could introduce continuous strategy variables to characterize gradual adjustments of subjects' strategies during data sharing. Secondly, this article verifies the effectiveness of the model through numerical simulation but does not conduct empirical testing based on actual research data. Subsequently, an empirical section can be conducted through research questionnaires and case studies to demonstrate the effectiveness of the model.

Author Contributions: Conceptualization, Z.Z., L.X., Y.S. (Yan Shi) and Y.S. (Yongqiang Shi); methodology, Z.Z., L.X., Y.S. (Yan Shi) and Y.S. (Yongqiang Shi); software, L.X.; validation, L.X. and Y.S. (Yan Shi); formal analysis, L.X. and Y.S. (Yongqiang Shi); investigation, L.X.; resources, Z.Z.; data curation, L.X.; writing—original draft preparation, L.X.; writing—review and editing, Z.Z., Y.S. (Yan Shi) and Y.S. (Yongqiang Shi); visualization, L.X.; supervision, Z.Z.; project administration, Y.S. (Yan Shi) and Y.S. (Yongqiang Shi); funding acquisition, Z.Z. and Y.S. (Yan Shi). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guangxi Natural Science Foundation Program, grant number 2023GXNSFAA026308; the Fundamental Research Funds for the Central Universities; and the National Natural Science Foundation of China (NSFC), grant number 72461004.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data and information within this manuscript are in the form of tables and other details.

Acknowledgments: The authors would like to express their sincere acknowledgement to the reviewers and editors for their time and efforts in providing comments and suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. General Office of the State Council. Opinions of the General Office of the State Council on Developing the Silver-Hair Economy and Enhancing the Well-Being of the Elderly. Available online: https://www.gov.cn/zhengce/content/202401/content_6926087.htm (accessed on 10 October 2025).
2. Chen, C.; Ran, X.; Liu, F.; Li, Y.; Li, N.; Zhang, Q. Reflections and Suggestions on Developing the Silver Economy to Cope with Population Aging. *Strateg. Study CAE* **2024**, *6*, 10–18. [CrossRef]
3. Zhang, X.; Liu, S.; Chen, X.; Wang, L.; Gao, B.; Zhu, Q. Health Information Privacy Concerns, Antecedents, and Information Disclosure Intention in Online Health Communities. *Inf. Manag.* **2018**, *55*, 482–493. [CrossRef]
4. Sun, Q.; Hao, G.; Ding, Y. Digital Platform Drives Supply Structure Optimization of Smart Elderly Care Services: Game Analysis Based on Perspective of Value Co-creation Theory. *Commer. Res.* **2023**, *65*, 58–69.
5. Chen, S.; Li, J.; Zhao, Y.; Zhu, Q. Exploration of Digital Intelligence Transformation of Smart Senior Care Service under the Data Factor Multiplier Effect. *Inf. Study Theory Appl.* **2025**, *48*, 1–9.
6. Ma, Y. Promote integrated development of digital economy and silver economy. *Macroecon. Manag.* **2022**, *3*, 56–62. [CrossRef]
7. Albarqi, M.N. Exploring the Effectiveness of Technology-Assisted Interventions for Promoting Independence in Elderly Patients: A Systematic Review. *Healthcare* **2024**, *12*, 2105. [CrossRef]
8. Grimm, M.; Link, E.; Albrecht, M.; Czerwinski, F.; Baumann, E.; Suhr, R. Exploring Functions and Predictors of Digital Health Engagement Among German Interest Users: Survey Study. *J. Med. Internet Res.* **2023**, *25*, e44024. [CrossRef]
9. Lyles, C.R.; Oanh, K.N.; Khoong, E.C.; Aguilera, A.; Sarkar, U. Multilevel Determinants of Digital Health Equity: A Literature Synthesis to Advance the Field. *Annu. Rev. Public Health* **2024**, *44*, 383–405. [CrossRef]
10. Wu, Y.Z. *Developing the Silver Economy to Cope with an Aging Population*; Fudan Development Institute: Shanghai, China, 2024; Available online: https://fddi.fudan.edu.cn/_t2515/7e/81/c21257a687745/page.htm (accessed on 17 October 2025).
11. Peng, Z.; Chen, Q. On China's Silver economy. *Chin. Soc. Secur. Rev.* **2022**, *6*, 49–66.
12. Xu, Y.; Liu, H. Current situation, problems and prospects of silver economy in China. *J. Beijing Univ. Aeronaut. Astronaut. (Soc. Sci. Ed.)* **2023**, *36*, 140–147.
13. Jin, N.; Yuan, X. High quality development of silver economy: Population foundation, strategic direction and path selection. *Hebei Acad. J.* **2024**, *44*, 158–166.
14. Zeng, H.; Fan, X. Further stimulate the silver market. *Macroecon. Manag.* **2019**, *10*, 33–38. [CrossRef]
15. Shu, Y.; Huang, J. Comparison of the Silver Economy between China and Asia-Pacific Developed Countries. *Asia-Pac. Econ. Rev.* **2023**, *1*, 91–97. [CrossRef]
16. Li, X.; Xie, J.; Luo, J.; Yang, A. The Silver-Hair Economy in the New Era: Political Economy Perspectives on Its Dilemmas and Solutions. *Sustainability* **2025**, *17*, 6760. [CrossRef]
17. Meng, Y.; Yang, H.C. The Mechanism and Practical Pathways of Fiscal and Tax Policies in Promoting High-quality Development of the Silver Economy. *Econ. Probl.* **2025**, *6*, 108–117.
18. Zhu, H.; Li, F. The Key Direction and Countermeasures of China's Silver Economic Development During the Period of the 15th Five-Year Plan (2026–2030). *J. Jiangxi Norm. Univ. (Philos. Soc. Sci. Ed.)* **2025**, *58*, 37–48.
19. He, Y.; Li, K.; Wang, Y. Crossing the digital divide: The impact of the digital economy on elderly individuals' consumption upgrade in China. *Technol. Soc.* **2022**, *71*, 102141. [CrossRef]
20. Guan, J.; Xu, H.; Wang, Y.; Ma, Y.; Wang, Y.; Gao, R.; Yu, K. Digital Economy and Health: A Case Study of a Leading Enterprise's Value Mining Mode in the Global Big Health Market. *Front. Public Health* **2022**, *10*, 904186. [CrossRef]
21. Fan, C.; Mao, Q.; Sun, J. Does digital financial inclusion promote elderly consumption? An empirical test based on Chinese survey data. *Electron. Commer. Res.* **2024**. [CrossRef]

22. Zeng, Y.; Fu, J.; Gao, W.; Chen, Y.; Yang, Z. Research Hotpots and Knowledge Framework of Digital Healthcare Service Management from the Perspective of the Silver Economy. *Sustainability* **2024**, *16*, 9735. [CrossRef]
23. Liu, L.; Lyu, H.; Dai, J.; Tu, Y.; Gao, T. Data-Driven Spatial Optimization of Elderly Care Facilities: A Study on Nonlinear Threshold Effects Based on XGBoost and SHAP—A Case Study of Xi'an, China. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 371. [CrossRef]
24. Li, C.; Li, X.; Wang, Y. Evolutionary Game Analysis of the Supervision Behavior for Public-Private Partnership Projects with Public Participation. *Math. Probl. Eng.* **2016**, *2016*, 1760837. [CrossRef]
25. Zhu, G.; Liu, H.; Feng, M. An Evolutionary Game-Theoretic Approach for Assessing Privacy Protection in mHealth Systems. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2196. [CrossRef] [PubMed]
26. Ho, E.; Rajagopalan, A.; Skvortsov, A.; Arulampalam, S.; Piraveenan, M. Game Theory in Defence Applications: A Review. *Sensors* **2022**, *22*, 1032. [CrossRef]
27. Duan, W.; Li, C.; Zhang, P.; Chang, Q. Game modeling and policy research on the system dynamics-based tripartite evolution for governmental regulation. *Clust. Comput.* **2016**, *19*, 2061–2074. [CrossRef]
28. Colman, A.M. Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behav. Brain Sci.* **2003**, *20*, 139–153. [CrossRef]
29. Shi, Q.; Yang, S.; Wang, N.; Zhang, S.; Wang, Y.; Wu, B.; Lu, X.; She, Y.; Yue, Z.; Gao, L.; et al. An evolutionary game-based simulation study of a multiagent governance system for smart senior care services in China. *BMC Geriatr.* **2023**, *23*, 871. [CrossRef]
30. Wang, Q.; Liu, J.; Zheng, Y. Evolutionary game and stability analysis of elderly care service quality supervision from the perspective of government governance. *Front. Public Health* **2023**, *11*, 1218301. [CrossRef]
31. Zhang, Z.; Song, X.; Shi, Y. Evolutionary Game Analysis of Behavior Strategies of Multiple Stakeholders in an Elderly Care Service System. *Int. J. Environ. Res. Public Health* **2023**, *20*, 4263. [CrossRef]
32. Wang, Q.; Liu, J.; Zheng, Y. Evolutionary game analysis of community elderly care service regulation in the context of "Internet+". *Front. Public Health* **2022**, *10*, 1093451. [CrossRef]
33. Zhang, L.; Lv, N.; Chen, N. Evolutionary Game of Medical Knowledge Sharing Among Chinese Hospitals Under Government Regulation. *Systems* **2025**, *13*, 454. [CrossRef]
34. Zhang, M.; Mu, D.; Deng, J.; Wang, P.; Li, Y. Research on the Collaborative Mechanism of Medical Data Sharing Stakeholders: Based on Quadripartite Evolutionary Game Analysis. *J. China Soc. Sci. Tech. Inf.* **2025**, *44*, 61–74.
35. Li, C.; Liu, C.; Ju, W. Dilemma and sustainable development strategy of sharing teaching resources for big data majors in universities based on game theory. *Discov. Artif. Intell.* **2024**, *4*, 103. [CrossRef]
36. Dong, Y.; Sun, Z.; Qiu, L. A Tripartite Evolutionary Game Analysis of Enterprise Data Sharing Under Government Regulations. *Systems* **2025**, *13*, 151. [CrossRef]
37. Annabelle, G. Bridging differing perspectives on technological platforms: Toward an integrative framework. *Res. Policy* **2014**, *43*, 1239–1249. [CrossRef]
38. Lu, Y.; Tang, Y. Interactive Service Model of Providing Services for the Aged in the Era of Digital Governance: Characteristics, Problems and Countermeasures. *Nanjing J. Soc. Sci.* **2020**, *7*, 51–59.
39. Sui, D.; Cao, X. The Selection and Optimization of the Supply Model of the "Internet+"–Pension Platform Based on Dynamic and Static Game Analysis. *J. Shaanxi Norm. Univ. (Philos. Soc. Sci. Ed.)* **2018**, *47*, 74–82.
40. Jia, Y.; Lan, Z.; Liu, R. Exact Senior Care: A New Senior Care Model Driven by Big Data. *J. Public Manag.* **2020**, *17*, 95–103.
41. Grover, V.; Chiang, R.H.; Liang, T.P.; Zhang, D. Creating strategic business value from big data analytics: A research framework. *J. Manag. Inf. Syst.* **2018**, *35*, 388–423. [CrossRef]
42. Ma, Y.; Li, Z.; He, L. Data Elements Marketization and Corporate Investment Efficiency: Causal Inference via Double Machine Learning. *Systems* **2025**, *13*, 609. [CrossRef]
43. Friedman, D. On Economic Applications of Evolutionary Game Theory. *J. Evol. Econ.* **1998**, *8*, 15–43. [CrossRef]
44. Taylor, P.D.; Jonker, L.B. Evolutionary Stable Strategies and Game Dynamics. *Math. Biosci.* **1978**, *40*, 145–146. [CrossRef]
45. Dong, C.; Liu, J.; Mi, J. How to Enhance Data Sharing in Digital Government Construction: A Tripartite Stochastic Evolutionary Game Approach. *Systems* **2023**, *11*, 212. [CrossRef]
46. Weibull, J.W. *Evolutionary Game Theory*; The MIT Press: Cambridge, MA, USA, 1997.
47. Dragicevic, A.Z. Reflective Evolution Under Strategic Uncertainty. *Int. J. Bifurc. Chaos Appl. Sci. Eng.* **2019**, *29*, 1950018. [CrossRef]
48. Mou, S.; Zhong, K.; Ma, Y. Regulating the Big Data-Based Discriminatory Pricing in Platform Retailing: A Tripartite Evolutionary Game Theory Analysis. *Mathematics* **2023**, *11*, 2579. [CrossRef]
49. Petrosyan, L.A.; Liu, X. Evolutionary Stable Strategies in Multistage Games. *Mathematics* **2023**, *11*, 2492. [CrossRef]
50. Ritzberger, K.; Weibull, J.W. Evolutionary selection in formal-form games. *Econometrica* **1995**, *63*, 1371–1399. [CrossRef]
51. Selten, R. A note on evolutionary stable strategies in asymmetric animal conflicts. *J. Theor. Biol.* **1980**, *84*, 93–101. [CrossRef]
52. Kalman, R.; Bertram, J. Control system analysis and design via the second method of Lyapunov: (I) continuous-time systems (II) discrete time systems. *IRE Trans. Autom. Control* **1959**, *4*, 112. [CrossRef]
53. Gilbert, N.; Troitzsch, K. *Simulation for the Social Scientist*, 2nd ed.; Open University Press: Maidenhead, UK, 2005.

54. Xu, J.; Zhao, H.; Han, H. *Annual Report on the Development of Silver Economy in China (2024)*; Social Sciences Academic Press: Beijing, China, 2024.
55. Li, D.; Mei, X. Evolutionary Game analysis of data sharing among large and medium-sized enterprises in the perspective of platform empowerment. *Sci. Rep.* **2024**, *14*, 11447. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Platform Data Transaction and Government Regulation Under Users' Personal Data Use Authorization Mechanism

Zhiwen Li *, Jing Li, Miao Qi and Zhiwei Chen

School of Management, Jiangsu University, Zhenjiang 212001, China; jingli@stmail.ujs.edu.cn (J.L.); miaoqi@stmail.ujs.edu.cn (M.Q.); zhiweichen@stmail.ujs.edu.cn (Z.C.)
* Correspondence: zhiwenli@ujs.edu.cn

Abstract

With the significant increase in users' privacy awareness and strict regulatory response, user privacy protection faces a dilemma between platform operational flexibility and commercial potential. The personal data use authorization has become an important mechanism influencing platform data transactions and government regulation, thereby shaping the development trajectory of the data trading market. Previous literature has yet to fully elucidate the complex interactions and evolutionary equilibrium strategies among the parties involved in data transactions. This study builds an evolutionary game model involving three parties—users, platforms, and the government—to explore the evolutionary equilibrium of users' personal data use authorization strategy, platforms' data trading strategy, and the government's regulatory strategy. The findings reveal that: (1) User authorization does not always promote platform data transactions. Under the personal data use authorization mechanism, user authorization facilitates platform data transactions only when either the data transaction price is low and the value of user behavioral data exceeds a certain threshold, or the data transaction price is high and the value of user behavioral data falls within a specific range. (2) User authorization is not a sufficient condition for the government to impose strict regulatory measures on platform data transactions. When the regulatory cost is low, the government's strict regulatory strategy is influenced by both the data transaction price and the value of user behavioral data. Specifically, under low-price–medium-value or high-price–high-value conditions, the government tends to strictly regulate platform data transactions when users do not authorize the use of personal data. Conversely, under low-price–high-value or high-price–medium-value conditions, the government is more inclined to enforce strict regulation when users authorize the use of personal data. This study contributes to evaluating the role of the personal data use authorization mechanism, offering valuable insights to platforms and the government on data transactions and regulation.

Keywords: evolutionary game; user authorization; data trading; personal data; user behavioral data

1. Introduction

As with most two-sided markets, platforms continuously accumulate data resources by connecting different user groups. These data resources help platforms gain deep insights into users' needs and accurately predict market trends, as well as driving service optimization and product iteration. Currently, more and more platforms' development strategies are evolving from "economies of scale" to "economies of scope", with expanding

into peripheral areas while consolidating their core markets. The demand for business diversification has directly led to an increase in platform data transactions. Data transactions refer to platforms exchanging or acquiring each other's market user data based on data sharing or trading mechanisms. For example, Meituan and Baidu Maps, two leading Chinese enterprises in the local lifestyle services and online map services markets, respectively, are achieving mutual benefits through data transactions and sharing. Specifically, Meituan provides Baidu Maps with comprehensive data on local lifestyle service providers, including restaurants and entertainment venues, as well as user review data. This significantly enriches the map information and service content of Baidu Maps, enhancing its service quality and user experience. In return, Baidu Maps provides Meituan with precise geographical location and traffic flow data, which are crucial for optimizing logistics delivery routes and improving the accuracy of merchant recommendation algorithms. This model of platform data transactions enables the mutual exchange of resources among platforms, promoting the efficient utilization of data resources and fostering the synergistic development of inter-platform businesses. This injects new vitality into the prosperity of the digital economy.

However, obtaining user authorization for personal data is crucial during platform data transactions. As users' awareness of data privacy protection has increased significantly, they are demanding greater control over their personal data and expecting transparency and informed consent during the collection and use of their data. China has recently issued several policy documents on data management, including the "Data Security Law of the People's Republic of China", the "Personal Information Protection Law of the People's Republic of China", and the "Electronic Commerce Law of the People's Republic of China". These regulations establish a normative framework for the collection, storage and transactions of data. They require platforms to strictly comply with legal provisions when conducting data transactions, strengthen user privacy protection and ensure the legality of data. Against this backdrop, platforms typically require users to read and agree to the data usage terms and conditions during registration, clarifying the scope of data use and potential sharing with third parties. Particularly, platforms commonly implement personal data authorization mechanisms, enabling users to access privacy settings and specify who can access their personal information. These privacy settings balance user authorization and data utilization, enabling users to control and manage their personal data. Under these mechanisms, platforms must obtain prior authorization from users when intending to trade or share their personal data.

While these mechanisms are effective in protecting user privacy and data security, they can also constrain platforms' operational flexibility and limit their commercial potential, thereby undermining their capacity for service innovation and market expansion. In this context, exploring how user authorization impacts platform data transactions, thereby enriching data products and services and attracting more entities to participate in data transactions, is of great significance for promoting the development of the data transaction market. However, despite the growing discussion of data rights, existing research lacks a systematic theoretical framework with which to analyze the interactive mechanisms between the micro-level behaviors of users authorizing transactions and the transactions themselves. This hinders the provision of accurate assessments of the practical utility of user data ownership in market-based data allocation. At the regulatory policy level, an effective link between data ownership classification and differentiated regulatory strategies has yet to be established. Existing regulatory theories appear to be insufficiently adaptive when addressing complex ownership forms, such as the integration of personal and behavioral data. Additionally, similar studies on topics such as information/data protection policies and their impact on data transactions have primarily relied on static models, such

https://doi.org/10.3390/jtaer21010012

as the Hotelling model, for analysis [1]. These models have yet to fully unravel the complex interactions and evolutionary equilibrium strategies among the parties involved. This study therefore examines users, platforms and the government operating under bounded rationality, whose strategic choices involve balancing dynamic data value against costs. The study focuses on two primary questions: (1) Under what specific conditions does user authorization facilitate or hinder platform data transactions? (2) How should government regulatory strategies be dynamically adjusted in response to the interplay of user authorization and platform decision? This study contributes by applying an evolutionary game framework to construct a tripartite model involving users, platforms, and the government in the context of data transactions. It analyses the dynamic evolution of user personal data authorization strategy, platform data transaction strategy and government regulatory strategy, exploring the evolutionary logic of tripartite strategy selection under user authorization mechanisms.

2. Literature Review

In recent years, data transactions and regulatory issues on platforms have become a hot topic of academic research. This study reviews the related literature primarily from the perspectives of data characteristics and value, data transactions, and the regulation of data transactions.

2.1. Research on Data Characteristics and Value

"Data" can be regarded as the portion of "information" that does not constitute "thought" or "knowledge" [2]. The same dataset can be used simultaneously by multiple enterprises or individuals [3]. Data possesses characteristics such as non-rivalry [4], non-excludability [5], uncertainty regarding data rights [6] and network externalities [7]. There is no absolute optimal solution regarding whether individuals or enterprises should hold data ownership, as this depends on several key factors [8,9]. The significant value of data manifests in several aspects: First, the development and application of data have shifted corporate decision-making from experience-driven to "data-driven decision-making," substantially enhancing corporate revenue and competitive advantage [10–13]. Second, data boosts productivity and efficiency. Leveraging data for innovation facilitates the production of final goods in key economic sectors, e.g., agriculture [14,15], thereby driving economic growth [16]. Supported by digital technologies, modern enterprises can access vast amounts of data and utilize it to pursue various competitive advantages [17]. Third, data influences platform pricing behavior [18], enabling decision-makers to design personalized pricing algorithms based on existing data [19–21].

2.2. Research on Data Transactions

Current research on data transactions primarily focuses on the following aspects: First, facilitating data transactions is a crucial step in establishing and improving data factor markets while unlocking the potential of data as a production factor. Data sharing leverages the non-rivalrous nature of data, enabling more enterprises to utilize open data for technological research, development, and innovation [22]. Transaction markets promote the circulation of innovative ideas, enhance corporate productivity, and stimulate economic growth [23]. Second, regarding data transactions and pricing, data transaction pricing is influenced by the consistency between individual privacy considerations and rational preferences [24]. Pricing data elements requires specific transaction contexts and necessitates targeted valuation of data element value based on typical application scenarios [25–27]. Depending on the situation, this can be categorized as traditional accounting pricing [28], or multidimensional pricing of data asset value [29]. Data transactions and welfare can

https://doi.org/10.3390/jtaer21010012

be maximized through logarithmic pricing functions [30]. Third, data markets play a crucial role in facilitating data circulation [31]. As data trading evolves, several challenges emerge: (1) Data-holding companies resist sharing data [32,33]. Despite the benefits of data circulation, firms often conceal data to maintain competitive advantages, creating data silos that severely hinder the realization of its economic value [2]. (2) Data ownership remains ambiguous. Johnson et al. note that strong property rights may lead to insufficient data sharing, harming consumers and advertisers, and propose reallocating property rights to consumers [34]. (3) Balancing "data security" and "data circulation" poses challenges. The negative externalities of data privacy may trigger "excessive data sharing," creating a paradox where both data security and circulation cannot be simultaneously achieved in data transactions [35–37]. Sharing may compromise "data security", highlighting the potential conflict between "data security" and "data circulation" [38–40], necessitating a balance between the two [41]. (4) Information asymmetry exists during data transactions [42]. Sellers hold an informational advantage regarding data quality, sources, and compliance, while buyers possess greater clarity about their intended use and the data's ultimate value. This mutual uncertainty forms the tension in data transactions. In emerging data markets characterized by information asymmetry, trust is the key factor that facilitates transactional cooperation [43].

2.3. Research on the Regulation of Data Transactions

On one hand, some scholars have emphasized the necessity of data transaction regulation. Wang argues that the effectiveness of government data privacy legislation and the brand image of online enterprises positively influence trust in online businesses [44]. Sun et al. contend that personal data has become a vital resource for platform enterprises, discussing the potential economic impact of government data regulation on platforms and the role of personal data in fostering value creation and innovation [45]. Kang et al. contend that privacy regulation can increase users' willingness to share personal data by strengthening their psychological ownership of such data and alleviating privacy concerns [46]. Marikyan argues that legal mechanisms must be introduced to safeguard the privacy and security of data exchanged between individuals and organizations [47]. On the other hand, some scholars have pointed out the possible negative effects caused by data transaction regulation. For example, data regulations in China, Europe, and the United States restrict internet platforms' ability to collect and use personal data for personalized recommendation and may fundamentally impact internet commerce [45].

2.4. Identified Gaps and Research Contributions

A systematic review of extant literature reveals the following areas that merit attention to enhance the quality of current research: Firstly, despite the increasing academic exploration of data property rights, no systematic theoretical framework has yet been established to address the interactive relationship between user data authorization mechanisms and platform data transactions. The extant research has focused predominantly on the legal attributes of data rights confirmation, overlooking the impact of micro-level user authorization behaviors on platform data transactions. This oversight impedes the precise evaluation of the practical efficacy of user data ownership in the market-based allocation of data as a factor of production. Secondly, at the level of regulatory policy, extant literature predominantly accentuates the institutional functions of government regulation, yet it fails to establish a nexus between data ownership classifications and regulatory strategy selection. This limitation hinders the capacity of extant research to furnish an efficacious analytical framework for formulating diversified regulatory approaches. This is particularly evident in the context of addressing complex data ownership forms in the platform economy, such

as the blending of personal and behavioral data. Current regulatory theories demonstrate an inadequate degree of adaptability in such situations.

Based on related work on data ownership confirmation, such as the work by Mahajan [48] and Li [49], the present study proposes a tripartite evolutionary game model encompassing users, platforms, and the government. By analyzing the dynamic evolution of strategic equilibria among user authorization behaviors, platform data transaction decisions, and government regulatory measures under a user data authorization mechanism, this study explores how the interaction between user authorization behaviors and government regulatory policies shapes and influences platform data transaction decisions. The objective of this research is to establish the theoretical underpinnings for platforms and governments in the context of data transactions and their regulation.

3. Evolutionary Game Model Description and Assumptions

3.1. Evolutionary Game Model

Data transactions involve three key parties: the government, platforms, and users. Each party acts with bounded rationality and possesses its own strategy space. The game relationships among these three parties are illustrated in Figure 1.

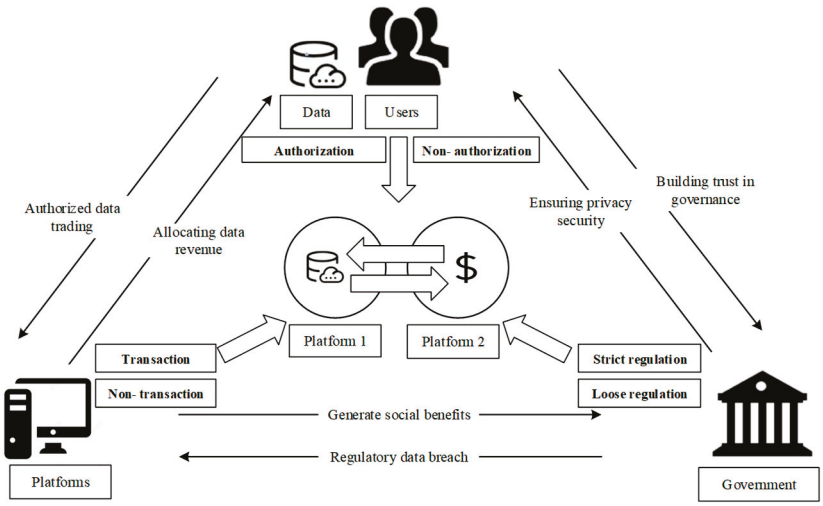


Figure 1. Tripartite game involving users, platforms, and the government.

For government, two regulatory strategies are available: strict regulation or loose regulation. Under strict regulation, platform data transactions are effectively regulated and supervised, ensuring user data security. However, the government must bear the associated regulatory costs. In the presence of loose regulation, transactions involving platform data are prone to inadequate regulation, which can lead to significant risks, including data security breaches. As a result, the government is often compelled to bear the consequences of reputational damage caused by data leaks. For platforms, two strategic options are available: the first is to engage in data transactions, and the second is to not engage in such transactions. Engaging in data transactions enables platforms to generate substantial profits by selling user data or data analysis services, thereby improving products and services. However, engaging in data transactions entails the risk of data leaks post-transaction. Not engaging in data transactions, on the one hand, serves to mitigate the risk of user privacy data leaks, thereby enhancing the platform's brand image. However, this practice also hinders the extraction and realization of data value. Two strategies are available to users: they may authorize or not authorize platforms to use their personal data. When users authorize such access, they potentially risk data breaches but in return gain access to customized product and service recommendations and a streamlined consumption experience.

In instances where users do not authorize, their privacy and personal information are effectively protected; however, their service experience is compromised.

3.2. Assumptions

Assumptions 1. *All three parties in the evolutionary game possess bounded rationality. The strategic space of users is {Authorization, Non-authorization}, with respective probabilities of $x(0 \leq x \leq 1)$, $1-x$; the strategic space of platform is {Transaction, Non-transaction}, with probabilities of $y(0 \leq y \leq 1)$, $1-y$; the strategic space of government is {Strict regulation, Loose regulation}, with probabilities of $z(0 \leq z \leq 1)$, $1-z$.*

Assumptions 2. *Platform data resources comprise two categories: user personal data and user behavioral data [50]. User personal data includes individual information, browsing history, and bookmarked products or services. This data is individual-specific and can be used for precision marketing. User behavioral data is derived from processing and analyzing aggregate user behavior records, such as consumer psychological states, service preferences, and market trends. This data reflects general behavioral patterns of users, possesses group characteristics, and can be utilized for business opportunity forecasting and algorithm refinement. Both data types first generate value within the platform, termed internal value, denoted as V_0 and ΔV, respectively. This value manifests primarily through internal applications such as optimizing products and services, enhancing user experience, and improving operational efficiency. Greater internal value from these data resources yields increased benefits for both the platform and users. Let α and β represent the benefit coefficients derived by users and the platform, respectively, from the internal value of data resources.*

Assumption 3. *A platform may choose to trade its accumulated data resources with other platforms, thereby generating data transaction value. Here, γ represents the price coefficient for data transactions, and C represents the cost of data transactions. When users grant authorization, platforms can trade both personal data and behavioral data resources together, with a data leakage probability of ε_1. When users do not grant authorization, platforms can only trade behavioral data, with a data leakage probability of ε_2. Since data transfers and two-party storage are involved in platform data transactions when users grant authorization, the risk of data leakage is higher, so there is $\varepsilon_1 > \varepsilon_2$. Once data leakage occurs, the loss of user is set at L.*

Assumption 4. *From the societal perspective of developing and utilizing the value of data, the government can directly or indirectly derive benefits from platform data transactions. Specifically, when both user personal data and behavioral data are traded by platforms, the government benefit is w_1; when only behavioral data is traded by platforms, the government benefit is w_2. Under strict government regulation, regulatory costs amount to C_g. Users gain trust benefits of T_g from enhanced data security. Simultaneously, if data breaches occur during platform transactions, government fines imposed on platforms total f. Under loose regulation, the government incurs a reputational damage of F upon data breaches.*

Parameter symbols and their specific meanings are detailed in Table 1.

Table 1. Model symbols and their meanings.

Variable	Description	Range
α	User benefit coefficient derived from the value of data resources	$\alpha \in [0,1]$
β	Platform benefit coefficient derived from the value of data resources	$\beta \in [0,1]$
V_0	The value of user personal data	$V_0 \geq 0$
ΔV	The value of user behavior data	$\Delta V \geq 0$

https://doi.org/10.3390/jtaer21010012

Table 1. *Cont.*

Variable	Description	Range
γ	Price coefficient for platform data transactions	$\gamma \geq 0$
C	The cost of platform data transactions	$C \geq 0$
C_g	The cost of strict government regulation	$C_g \geq 0$
L	Financial losses for users caused by data breaches	$L \geq 0$
ε_1	Probability of data leakage when users authorize the platform to use their personal data	$\varepsilon_1 \geq 0$
ε_2	Probability of data leakage when users do not authorize the platform to use their personal data	$\varepsilon_2 \geq 0$
f	Government fines imposed on platforms following data breaches	$f \geq 0$
F	The reputational damage to the government caused by data breaches under loose regulation	$F \geq 0$
T_g	User trust benefit under strict government regulation	$T_g \geq 0$
w_1	Government revenue when both user personal data and behavioral data are traded	$w_1 \geq 0$
w_2	Government revenue when only user behavioral data is traded	$w_2 \geq 0$

3.3. Establishing the Revenue Matrix

Based on the above assumptions, the payoff matrices for each party in the three-party strategy evolution model involving the government, platforms, and users are shown in Table 2.

Table 2. Evolutionary game payoff matrix.

Users	Platforms	Government	
		Strict Regulation (z)	**Loose Regulation** $(1-z)$
Authorization (x)	Transaction (y)	$\alpha(V_0+\Delta V+\gamma(V_0+\Delta V))-\varepsilon_1 L+T_g$ $\beta(V_0+\Delta V+\gamma(V_0+\Delta V))-\varepsilon_1 f-C$ $w_1-C_g+\varepsilon_1 f$	$\alpha(V_0+\Delta V+\gamma(V_0+\Delta V))-\varepsilon_1 L$ $\beta(V_0+\Delta V+\gamma(V_0+\Delta V))-C$ $w_1-\varepsilon_1 F$
	Non-transaction $(1-y)$	$\alpha(V_0+\Delta V)+T_g$ $\beta(V_0+\Delta V)$ $-C_g$	$\alpha(V_0+\Delta V)$ $\beta(V_0+\Delta V)$ 0
Non-authorization $(1-x)$	Transaction (y)	$\alpha(\Delta V+\gamma\Delta V)-\varepsilon_2 L+T_g$ $\beta(\Delta V+\gamma\Delta V)-\varepsilon_2 f-C$ $w_2-C_g+\varepsilon_2 f$	$\alpha(\Delta V+\gamma\Delta V)-\varepsilon_2 L$ $\beta(\Delta V+\gamma\Delta V)-C$ $w_2-\varepsilon_2 F$
	Non-transaction $(1-y)$	$\alpha\Delta V+T_g$ $\beta\Delta V$ $-C_g$	$\alpha\Delta V$ $\beta\Delta V$ 0

4. Solving the Evolutionary Stabilization Strategy

4.1. Expected Payoff of Users

The expected payoff for users choosing the authorization strategy is U_{11}, and the expected payoff for users choosing the non-authorization strategy is U_{12}. Users' expected payoff is U_1. Therefore,

$$\begin{aligned} U_{11} = {} & yz\big(\alpha(V_0+\Delta V+\gamma V_0+\Delta V)-\varepsilon_1 L+T_g\big)+ \\ & y(1-z)(\alpha(V_0+\Delta V+\gamma V_0+\Delta V)-\varepsilon_1 L)+ \\ & (1-y)z\big(\alpha(V_0+\Delta V)+T_g\big)+(1-y)(1-z)(\alpha(V_0+\Delta V)) \end{aligned} \tag{1}$$

$$U_{12}= yz(\alpha(1+\gamma)\Delta V-\varepsilon_2 L+Tg)+y(1-z)(\alpha(1+\gamma)\Delta V-\varepsilon_2 L)+ (1-y)z(\alpha\Delta V+T_g)+(1-y)(1-z)(\alpha\Delta V) \tag{2}$$

$$U_1 = x\times U_{11}+(1-x)\times U_{12} \tag{3}$$

The replication dynamic equation of users is:

$$f(x)=(1-x)x(y(\alpha(1-\gamma)\Delta V+L(\varepsilon_2-\varepsilon_1))+V_0\alpha(1+y\gamma)) \tag{4}$$

The first derivative of the replication dynamic equation is:

$$f'(x)=\frac{\partial f(x)}{\partial x}=(1-2x)(y(\alpha(1-\gamma)\Delta V+L(\varepsilon_2-\varepsilon_1))+V_0\alpha(1+y\gamma)) \tag{5}$$

According to the stability theorem of the replicator dynamics equation, to achieve policy stability, conditions $f(x)=0$ and $f'(x)<0$ must be satisfied. The corresponding point is then the stable point of the evolutionary game.

Letting $F(y)=y(\alpha(1-\gamma)\Delta V+L(\varepsilon_2-\varepsilon_1))+V_0\alpha(1+y\gamma)=0$ yields:

$$y=y^*=\frac{\alpha V_0}{(\varepsilon_1-\varepsilon_2)L-\alpha\gamma V_0-\alpha(1-\gamma)\Delta V} \tag{6}$$

Therefore, when $0<y<y^*$, there is $f'(x)|_{x=1}<0$, $f'(x)|_{x=0}>0$, and $x=1$ represents a stable point in the evolutionary game, at which users choose the authorization strategy; when $y^*<y<1$, there is $f'(x)|_{x=0}<0$, $f'(x)|_{x=1}>0$, and $x=0$ represents a stable point in the evolutionary game, at which users choose the non-authorization strategy; when $y=y^*$, there is $f(x)\equiv 0$, it indicates that all points on x are in a stable state, meaning users' strategy does not change over time. This leads to the following proposition:

Proposition 1. *When $0<y<y^*$, $x=1$ is users' evolutionarily stable strategy; when $y^*<y<1$, $x=0$ is users' evolutionarily stable strategy; when $y=y^*$, the stable strategy cannot be determined.*

Proposition 1 indicates that, with other parameters constant, users' strategy is influenced to some extent by platforms' initial strategy probability. As platforms' transaction probability increases, the probability of users' authorization strategy decreases. Figure 2 shows the phase diagram of users' strategy evolution dynamics.

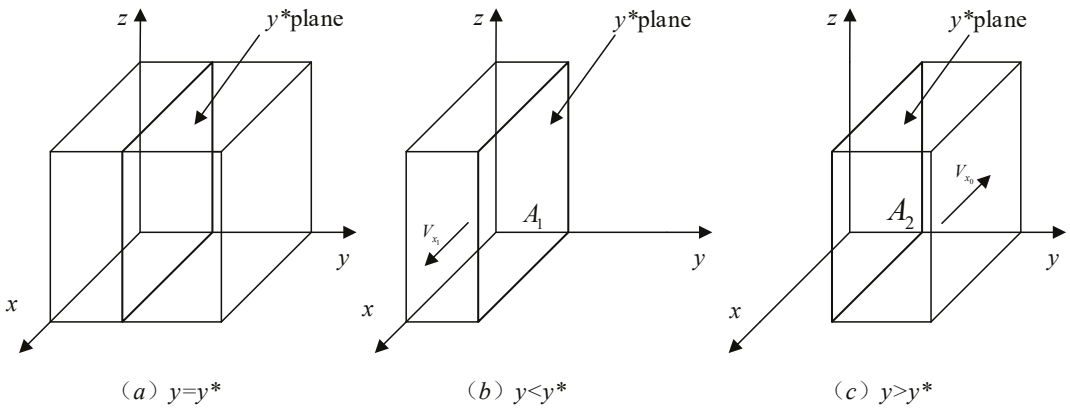


Figure 2. Dynamic phase diagram of users' strategy evolution.

The first-order partial derivatives of each parameter yield the following results. When $\Delta V>\frac{L(\varepsilon_1-\varepsilon_2)}{\alpha(1-\gamma)}$, there is $\frac{\partial y^*}{\partial V_0}>0$, indicating that $y*$ increases as V_0 increases; the $y*$ plane moves along the positive direction of the y axis. As a result, the volume of A_1, which is the probability of user authorization, increases. When $\Delta V<\frac{L(\varepsilon_1-\varepsilon_2)}{\alpha(1-\gamma)}$, there is $\frac{\partial y*}{\partial V_0}<0$,

indicating that $y*$ decreases as V_0 increases; the volume of A_1 decreases, indicating a reduced probability of authorization. Similarly, when $\gamma < 1$, there is $\frac{\partial y^*}{\partial \Delta V} > 0$; when $\gamma > 1$, there is $\frac{\partial y^*}{\partial \Delta V} < 0$; and $\frac{\partial y^*}{\partial \varepsilon_2} > 0$, $\frac{\partial y^*}{\partial \alpha} > 0$, $\frac{\partial y^*}{\partial \gamma} > 0$, $\frac{\partial y^*}{\partial \varepsilon_1} < 0$, $\frac{\partial y^*}{\partial L} < 0$. This leads to the following proposition:

Proposition 2. *Regarding the probability of users selecting the authorization strategy, when $\Delta V > \frac{L(\varepsilon_1-\varepsilon_2)}{\alpha(1-\gamma)}$, it is positively correlated with V_0; when $\gamma < 1$, it is positively correlated with ΔV. Additionally, it is positively correlated with ε_2, α, and γ, but is negatively correlated with ε_1 and L.*

Proposition 2 illustrates that users' authorization decisions involve a trade-off between benefits and risks. Specifically, the greater the benefit coefficient users derive from data value, or the higher the price coefficient of data transactions, the greater their potential gains—thereby increasing their inclination to authorize. Conversely, the lower the probability of data leakage when not authorizing, or the higher the probability of leakage when authorizing coupled with greater losses to user benefits from such leakage, the greater the risks of authorization—thereby increasing users' inclination to not authorize. Furthermore, only when the behavioral data value is high will users increasingly favor authorizing their personal data, since authorization allows them to gain more from platforms' data transactions. However, when the data transaction price is high, users will increasingly favor not authorizing their personal data as the value of their behavioral data increases. This is because they can already obtain sufficient benefits from platforms' behavioral data transactions, making them more inclined to not authorize.

4.2. Expected Payoff of Platforms

The expected payoff when platforms choose to engage in data transaction is U_{21}, the expected payoff when platforms choose not to engage in data transaction is U_{22}, and platforms' expected payoff is U_2. Therefore,

$$\begin{aligned} U_{21} = {} & xz(\beta(V_0 + \Delta V + \gamma V_0 + \Delta V) - \varepsilon_1 f - C) + \\ & x(1-z)(\beta(V_0 + \Delta V + \gamma V_0 + \Delta V) - C) + \\ & (1-x)z(\beta((1+\gamma)\Delta V) - \varepsilon_2 f - C) + \\ & (1-x)(1-z)(\beta((1+\gamma)\Delta V) - C) \end{aligned} \tag{7}$$

$$\begin{aligned} U_{22} = {} & xz\beta(V_0 + \Delta V) + x(1-z)\beta(V_0 + \Delta V) + \\ & (1-x)z\beta\Delta V + (1-x)(1-z)\beta\Delta V \end{aligned} \tag{8}$$

$$U_2 = y \times U_{21} + (1-y) \times U_{22} \tag{9}$$

The replication dynamic equation of platforms is:

$$f(y) = (1-y)y(\beta(x + (1-x)\gamma)\Delta V - z\varepsilon_2 f + xz(\varepsilon_2 - \varepsilon_1)f + x\beta\gamma V_0 - C) \tag{10}$$

The first derivative of the replication dynamic equation is:

$$f'(y) = (1-2y)(\beta(x + (1-x)\gamma)\Delta V - z(\varepsilon_2 + x(\varepsilon_2 - \varepsilon_1))f + x\beta\gamma V_0 - C) \tag{11}$$

Letting $F(x) = \beta(x + (1-x)\gamma)\Delta V - z(\varepsilon_2 + x(\varepsilon_2 - \varepsilon_1))f + x\beta\gamma V_0 - C = 0$ yields:

$$x = x^* = \frac{C - \beta\gamma\Delta V + z\varepsilon_2 f}{\beta(1-\gamma)\Delta V - z(\varepsilon_1 - \varepsilon_2)f + \beta\gamma V_0} \tag{12}$$

Therefore, when $0 < x < x^*$, there is $f'(y)|_{y=0} < 0$, $f'(y)|_{y=1} > 0$, and $y = 0$ is the equilibrium point of the evolutionary game, platforms adopt the non-transaction strategy at this point; when $x^* < x < 1$, there is $f'(y)|_{y=1} < 0$, $f'(y)|_{y=0} > 0$, and

https://doi.org/10.3390/jtaer21010012

$y = 1$ is the equilibrium point of the evolutionary game, at which platforms choose the transaction strategy; when $x = x^*$, there is $f(y) \equiv 0$, it indicates that all points on y are in a stable state, meaning platforms' strategy does not change over time. This leads to the following proposition:

Proposition 3. *When $0 < x < x^*$, $y = 0$ is platforms' evolutionarily stable strategy; when $x^* < x < 1$, $y = 1$ is platforms' evolutionarily stable strategy; when $x = x^*$, the stable strategy cannot be determined.*

Proposition 3 indicates that, holding other parameters constant, platforms' strategy is influenced to some extent by the probability of users' initial strategy. The higher the initial probability of users' authorization, the greater the probability that platforms will choose the transaction strategy. The phase diagram of platforms' strategy evolutionary dynamics is shown in Figure 3.

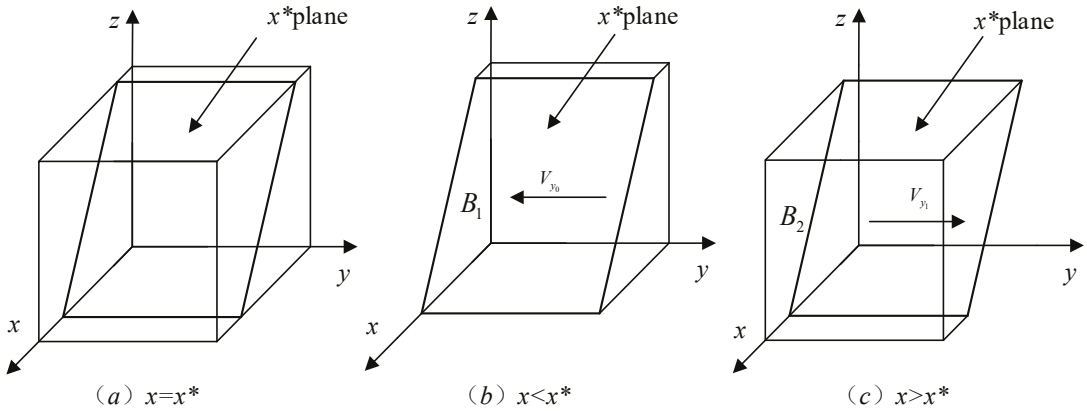


Figure 3. Dynamic phase diagram of platforms' strategy evolution.

Further analysis reveals that when $\Delta V > \frac{C+fz\varepsilon_2}{\beta\gamma}$, there is $\frac{\partial x^*}{\partial V_0} < 0$. It means that x^* decreases as V_0 increases, and plane x^* moves along the negative direction of axis x. As a result, the volume of B_2, which represents the probability of platforms selecting a transaction strategy, increases. Similarly, when $\gamma < \frac{C+fz\varepsilon_1-\sqrt{(C+fz\varepsilon_1)^2-4V_0\beta(C+fz\varepsilon_2)}}{2V_0\beta}$, there is $\frac{\partial x^*}{\partial \Delta V} < 0$; when $\gamma > \frac{C+fz\varepsilon_2}{C+fz\varepsilon_1}$ and $\Delta V > \frac{V_0\gamma(C+fz\varepsilon_2)}{C(\gamma-1)+fz(\gamma\varepsilon_1-\varepsilon_2)}$, there is $\frac{\partial x^*}{\partial \beta} < 0$; when $\Delta V > \frac{V_0(C+fz\varepsilon_2)}{C-\beta\Delta V+fz\varepsilon_1}$, there is $\frac{\partial x^*}{\partial \gamma} < 0$; when $\Delta V > \frac{C+fz\varepsilon_2}{\beta\gamma}$, there is $\frac{\partial x^*}{\partial \varepsilon_1} > 0$; when $\Delta V < \frac{C+fz\varepsilon_1-V_0\beta\gamma}{\beta}$, there is $\frac{\partial x^*}{\partial \varepsilon_2} > 0$; when $\gamma > \frac{\varepsilon_2}{\varepsilon_1}$ and $\Delta V > \frac{C(\varepsilon_1-\varepsilon_2)+V_0\beta\gamma\varepsilon_2}{\beta(\gamma\varepsilon_1-\varepsilon_2)}$, there is $\frac{\partial x^*}{\partial f} > 0$. The following proposition is thus summarized:

Proposition 4. *Regarding the probability of platforms selecting the transaction strategy, when $\Delta V > \frac{C+fz\varepsilon_2}{\beta\gamma}$, it is positively correlated with V_0 and negatively correlated with ε_1; When $\gamma < \frac{C+fz\varepsilon_1-\sqrt{(C+fz\varepsilon_1)^2-4V_0\beta(C+fz\varepsilon_2)}}{2V_0\beta}$, it is positively correlated with ΔV; When $\gamma > \frac{C+fz\varepsilon_2}{C+fz\varepsilon_1}$ and $\Delta V > \frac{V_0\gamma(C+fz\varepsilon_2)}{C(\gamma-1)+fz(\gamma\varepsilon_1-\varepsilon_2)}$, it is positively correlated with β; when $\Delta V > \frac{V_0(C+fz\varepsilon_2)}{C-\beta\Delta V+fz\varepsilon_1}$, it is positively correlated with γ; when $\Delta V < \frac{C+fz\varepsilon_1-V_0\beta\gamma}{\beta}$, it is negatively correlated with ε_2; when $\gamma > \frac{\varepsilon_2}{\varepsilon_1}$ and $\Delta V > \frac{C(\varepsilon_1-\varepsilon_2)+V_0\beta\gamma\varepsilon_2}{\beta(\gamma\varepsilon_1-\varepsilon_2)}$, it is negatively correlated with f.*

Proposition 4 indicates that while the probability of platforms' data transaction strategy is primarily determined by the relative strength of potential gains versus potential risks, it does not strictly increase with the growth of potential gains such as data value, benefit coefficient, or transaction price coefficient. Nor does it strictly decrease with the growth of

potential risks such as data leakage probability or government fines. These factors mutually constrain each other, collectively influencing the probability of platforms' data transaction strategy, resulting in a highly complex pattern of strategy selection.

4.3. Expected Payoff of Government

The expected payoff when the government chooses a strict regulation strategy is U_{31}, the expected payoff when the government chooses a loose regulation strategy is U_{32}, and the expected payoff of the government is U_3. Therefore,

$$\begin{aligned} U_{31} = & xy\left(w_1 - C_g + \varepsilon_1 f\right) - x(1-y)C_g + \\ & (1-x)y\left(w_2 - C_g + \varepsilon_2 f\right) - (1-x)(1-y)C_g \end{aligned} \tag{13}$$

$$U_{32} = xy(w_1 - \varepsilon_1 F) + (1-x)y(w_2 - \varepsilon_2 F) \tag{14}$$

$$U_3 = z \times U_{31} + (1-z) \times U_{32} \tag{15}$$

The replication dynamic equation of the government is:

$$f(z) = (1-z)z\left((f+F)y(x(\varepsilon_1 - \varepsilon_2) + \varepsilon_2) - C_g\right) \tag{16}$$

The first derivative of the replication dynamic equation is:

$$f'(z) = (1-2z)\left((f+F)y(x(\varepsilon_1 - \varepsilon_2) + \varepsilon_2) - C_g\right) \tag{17}$$

Letting $F(y) = (f+F)y(x(\varepsilon_1 - \varepsilon_2) + \varepsilon_2) - C_g = 0$ yields:

$$y = y^* = \frac{C_g}{(f+F)(x(\varepsilon_1 - \varepsilon_2) + \varepsilon_2)} \tag{18}$$

Therefore, when $0 < y < y^*$, there is $f'(z)|_{z=0} < 0$, $f'(z)|_{z=1} > 0$, and $z = 0$ is the equilibrium point of the evolutionary game, at which the government chooses the loose regulation strategy; when $y^* < y < 1$, there is $f'(z)|_{z=1} < 0$, $f'(z)|_{z=0} > 0$, and $z = 1$ is the equilibrium point of the evolutionary game, at which the government chooses the strict regulatory strategy; when $y = y^*$, there is $f(z) \equiv 0$, it indicates that all points on z are in a stable state, meaning the government's strategy does not change over time. This leads to the following proposition:

Proposition 5. *When $0 < y < y^*$, $z = 0$ is the government's evolutionarily stable strategy; when $y^* < y < 1$, $z = 1$ is the government's evolutionarily stable strategy; when $y = y^*$, the stable strategy cannot be determined.*

Proposition 5 indicates that, holding other parameters constant, the probability of the government choosing strict regulation is to some extent related to the probability of platforms' initial strategy. The higher the probability that platforms choose the transaction strategy, the greater the probability that the government will opt for strict regulation. The phase diagram of the government's strategy evolution dynamics is shown in Figure 4.

Computing the first-order partial derivatives of each parameter with respect to y^*, yields $\frac{\partial y^*}{\partial f} < 0$, $\frac{\partial y^*}{\partial F} < 0$, $\frac{\partial y^*}{\partial \varepsilon_1} < 0$ and $\frac{\partial y^*}{\partial \varepsilon_2} < 0$. That is to say, as f, F, ε_1, ε_2 increases, y^* decreases, the $y*$ plane moves toward the negative direction of axis y, and C_2 increases in volume—meaning the probability of the government choosing the strict regulatory strategy increases. Additionally, $\frac{\partial y^*}{\partial C_g} > 0$ suggests that y^* increases as C_g increases, the y^* plane moves toward the positive direction of axis y, and C_2 decreases in volume—meaning the

probability of the government choosing the strict regulation decreases. Based on the above, the following proposition is summarized:

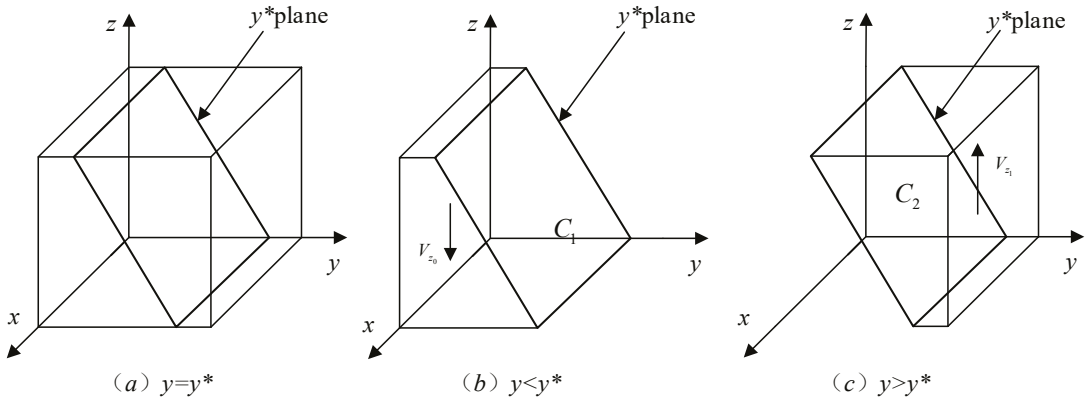


Figure 4. Dynamic phase diagram of the government's strategy evolution.

Proposition 6. *Regarding the probability of the government selecting the strict regulatory strategy, it is positively correlated with f, F, ε_1 and ε_2, but is negatively correlated with C_g.*

Proposition 6 indicates that government regulatory strategy is primarily based on a comparison between the benefits of strict regulation and the associated costs. When the fines imposed on platforms following data breaches, the probability of data leaks, and reputational damage are greater, the government tends to favor stricter regulation; conversely, when the costs of strict regulation are higher, it leans toward loose regulation.

4.4. Analysis of the Evolutionary Stabilization Strategy

Based on the above tripartite evolutionary game model and their respective expected payoffs, the three-dimensional replicating dynamical system equation is derived as follows:

$$\begin{cases} f(x) = (1-x)x(y(\alpha(1-\gamma)\Delta V - L(\varepsilon_1-\varepsilon_2)) + V_0\alpha(1+y\gamma)) \\ f(y) = (1-y)y(\beta(x+(1-x)\gamma)\Delta V + z(x(\varepsilon_2-\varepsilon_1)-\varepsilon_2)f + x\beta\gamma V_0 - C) \\ f(z) = (1-z)z\big(y(x(\varepsilon_1-\varepsilon_2)+\varepsilon_2)(f+F) - C_g\big) \end{cases} \tag{19}$$

Based on the three-dimensional replicating dynamic system model, the Jacobian matrix is obtained as follows:

$$J = \begin{bmatrix} (1-2x)\begin{pmatrix} \alpha(1+y\gamma)V_0 - \\ y(\alpha(\gamma-1)\Delta V + (\varepsilon_1-\varepsilon_2)L) \end{pmatrix} & (1-x)x\begin{pmatrix} \alpha\gamma V_0 - \alpha(\gamma-1)\Delta V - \\ L(\varepsilon_1-\varepsilon_2) \end{pmatrix} & 0 \\ (1-y)y\begin{pmatrix} \beta\gamma V_0 - \beta(\gamma-1)\Delta V \\ -fz(\varepsilon_1-\varepsilon_2) \end{pmatrix} & (1-2y)\begin{pmatrix} \beta(x+(1-x)\gamma)\Delta V - \\ z(\varepsilon_2+x(\varepsilon_2-\varepsilon_1))f + x\beta\gamma V_0 - C \end{pmatrix} & f(1-y)y(x(\varepsilon_2-\varepsilon_1)-\varepsilon_2) \\ (1-z)zy(\varepsilon_1-\varepsilon_2)(f+F) & (1-z)z(x(\varepsilon_1-\varepsilon_2)+\varepsilon_2)(f+F) & (1-2z)\big(y(x(\varepsilon_1-\varepsilon_2)+\varepsilon_2)(f+F)-C_g\big) \end{bmatrix} \tag{20}$$

Based on Lyapunov stability theory, the evolutionary game equilibrium in asymmetric games is a pure strategy equilibrium. Therefore, this study only investigates the pure strategy equilibrium points under conditions $f(x) = 0$, $f(y) = 0$, $f(z) = 0$. The system equilibrium points can be determined as $(0,0,0)$, $(0,1,0)$, $(0,1,1)$, $(0,0,1)$, $(1,0,0)$, $(1,1,0)$, $(1,1,1)$ and $(1,0,1)$. The system equilibrium points and eigenvalues are shown in Table 3.

Table 3. System equilibrium points and eigenvalues.

Equilibrium Point	λ_1	λ_2	λ_3
$(0,0,0)$	αV_0	$\beta\gamma\Delta V - C$	$-C_g$
$(0,1,0)$	$\alpha(1-\gamma)\Delta V + \alpha(1+\gamma)V_0 + (\varepsilon_2 - \varepsilon_1)L$	$C - \beta\gamma\Delta V$	$\varepsilon_2(f+F) - C_g$
$(0,1,1)$	$\alpha(1-\gamma)\Delta V + \alpha(1+\gamma)V_0 + (\varepsilon_2 - \varepsilon_1)L$	$C + \varepsilon_2 f - \beta\gamma\Delta V$	$C_g - \varepsilon_2(f+F)$
$(0,0,1)$	αV_0	$\beta\gamma\Delta V - \varepsilon_2 f - C$	C_g
$(1,0,0)$	$-\alpha V_0$	$\beta\Delta V + \beta\gamma V_0 - C$	$-C_g$
$(1,1,0)$	$(\varepsilon_1 - \varepsilon_2)L - \alpha(1+\gamma)V_0 - \alpha(1-\gamma)\Delta V$	$C - \beta\Delta V - \beta\gamma V_0$	$\varepsilon_1(f+F) - C_g$
$(1,1,1)$	$(\varepsilon_1 - \varepsilon_2)L - \alpha(1+\gamma)V_0 - \alpha(1-\gamma)\Delta V$	$C + \varepsilon_1 f - \beta\gamma V_0 - \beta\Delta V$	$C_g - \varepsilon_1(f+F)$
$(1,0,1)$	$-\alpha V_0$	$\beta\Delta V + \beta\gamma V_0 - f\varepsilon_1 - C$	C_g

4.5. Analysis of Stabilization Conditions

The equilibrium point is an evolutionarily stable strategy (ESS) if and only if the eigenvalues of its Jacobian matrix simultaneously meet $\lambda_1 < 0$, $\lambda_2 < 0$, and $\lambda_3 < 0$. Based on the model assumptions, we have: $\alpha V_0 > 0$, $C_g > 0$. Therefore, the only possible stable points are $(0,1,0)$, $(0,1,1)$, $(1,0,0)$, $(1,1,0)$, and $(1,1,1)$. Analyzing the eigenvalue conditions in Table 3 yields the following five scenarios.

Scenario 1. In this scenario, the evolutionary game's stable point is $(0,1,0)$, indicating the strategy combination of the three parties is (non-authorization, transaction, loose regulation). Specifically, users choose not to authorize, platforms opt for behavioral data transaction, and the government selects loose regulation. The conditions required for this stable point are in Table 4.

Table 4. The condition of γ, ΔV and C_g for the stable point $(0,1,0)$.

γ	ΔV	C_g
$0 < \gamma < 1$	$\frac{C}{\beta\gamma} < \Delta V < \frac{(\varepsilon_1-\varepsilon_2)L-\alpha(1+\gamma)V_0}{\alpha(1-\gamma)}$	$C_g > \varepsilon_2(f+F)$
$\gamma > 1$	$\Delta V > \max\left\{\frac{(\varepsilon_1-\varepsilon_2)L-\alpha(1+\gamma)V_0}{\alpha(1-\gamma)}, \frac{C}{\beta\gamma}\right\}$	$C_g > \varepsilon_2(f+F)$

This condition indicates that when the data transaction price is low and behavioral data value falls within a certain range, or when the data price is high and behavioral data value exceeds a specific threshold, users choose not to authorize the use of their personal data in exchange for greater benefits. Meanwhile, the profits generated by platforms trading behavioral data outweigh the costs of data transactions, prompting them to opt for data transaction. Simultaneously, the high costs of strict regulation render it unaffordable for government, leading it to choose loose regulation.

Scenario 2. In this scenario, the evolutionary game stabilizes at point $(0,1,1)$, indicating the strategy combination of the three parties is (non-authorization, transaction, strict regulation). Specifically, users choose not to authorize, platforms opt for behavioral data transaction, and the government opts for strict regulation. The conditions required for this stable point are in Table 5.

The conditions for this scenario are similar to those in Scenario 1. When behavioral data is valued within a certain range at a lower transaction price, or when its value exceeds a specific threshold at a higher transaction price, users choose not to authorize while platforms opt to transact. The key difference lies in the lower strict regulatory costs, which enable the government to achieve a balanced budget through revenues such as fines and

enhanced reputation. This creates a stronger incentive for the government to adopt stricter regulatory measures.

Table 5. The condition of γ, ΔV and C_g for the stable point $(0,1,1)$.

γ	ΔV	C_g
$0<\gamma<1$	$\frac{C+\varepsilon_2 f}{\beta\gamma}<\Delta V<\frac{(\varepsilon_1-\varepsilon_2)L-\alpha(1+\gamma)V_0}{\alpha(1-\gamma)}$	$C_g<\varepsilon_2(f+F)$
$\gamma>1$	$\Delta V>\max\left\{\frac{(\varepsilon_1-\varepsilon_2)L-\alpha(1+\gamma)V_0}{\alpha(1-\gamma)},\frac{C+\varepsilon_2 f}{\beta\gamma}\right\}$	$C_g<\varepsilon_2(f+F)$

Scenario 3. In this scenario, the evolutionary game's stable point is $(1,0,0)$, indicating the strategy combination of the three parties is (authorization, non-transaction, loose regulation). Specifically, users choose to authorize, platforms choose not to transact, and the government chooses loose regulation. The conditions required for this stable point are $\Delta V<\frac{C-\beta\gamma V_0}{\beta}$. In this scenario, the value of user behavioral data is relatively low. Users have an incentive to authorize platforms to use their personal data in exchange for greater benefits, leading them to choose authorization. However, the low value of behavioral data means platforms' revenue from data transactions is less than the cost of conducting such transactions, prompting platforms to choose not to trade. Platforms' non-trading strategy reduces the government's incentive for strict regulation, leading it to opt for loose regulation.

Scenario 4. In this scenario, the evolutionary game theory stable point is $(1,1,0)$. This stable point represents the strategy combination of the three parties: (authorization, transaction, loose regulation). Specifically, users choose authorization, platforms choose transaction, and the government chooses loose regulation. The conditions required for this stable point are in Table 6.

Table 6. The condition of γ, ΔV and C_g for the stable point $(1,1,0)$.

γ	ΔV	C_g
$0<\gamma<1$	$\Delta V>\max\left\{\frac{(\varepsilon_1-\varepsilon_2)L-\alpha(1+\gamma)V_0}{\alpha(1-\gamma)},\frac{C-\beta\gamma V_0}{\beta}\right\}$	$C_g>\varepsilon_1(f+F)$
$\gamma>1$	$\frac{C+\varepsilon_2 f}{\beta\gamma}<\Delta V<\frac{(\varepsilon_1-\varepsilon_2)L-\alpha(1+\gamma)V_0}{\alpha(1-\gamma)}$	$C_g>\varepsilon_1(f+F)$

This condition indicates that when the data transaction price is low and the value of user behavioral data exceeds a certain threshold, or when the data transaction price is high and the value of user behavioral data falls within a specific range, users can gain additional benefits by granting authorization, leading them to choose authorization. Simultaneously, platforms' revenue from trading personal data and behavioral data exceeds the transaction costs, prompting platforms to engage in data transaction. Significant strict regulatory costs diminish platforms' willingness to enforce strict regulation, compelling the government to adopt a more lenient regulatory strategy instead.

Scenario 5. In this scenario, the evolutionary game theory stable point is $(1,1,1)$. This stable point represents the strategy combination of the three parties: (authorization, transaction, strict regulation). Specifically, users choose authorization, platforms choose transaction, and the government chooses strict regulation. The conditions required for this stable point are in Table 7.

In this scenario, the strategy of users and platforms resembles that of Scenario 4. Users choose to authorize and platforms choose to engage in data transactions when behavioral data value exceeds a certain threshold under a lower data transaction price, or when

behavioral data value falls within a specific range under a higher data transaction price. The difference lies in the fact that larger fines and reputational gains enable the government to bear the costs of strict regulation, thereby incentivizing it to opt for strict regulation.

Table 7. The condition of γ, ΔV and C_g for the stable point $(1,1,1)$.

γ	ΔV	C_g
$0 < \gamma < 1$	$\Delta V > \max\left\{\frac{(\varepsilon_1-\varepsilon_2)L-\alpha(1+\gamma)V_0}{\alpha(1-\gamma)}, \frac{C+\varepsilon_1 f-\beta\gamma V_0}{\beta}\right\}$	$C_g < \varepsilon_1(f+F)$
$\gamma > 1$	$\frac{C+\varepsilon_1 f-\beta\gamma V_0}{\beta} < \Delta V < \frac{(\varepsilon_1-\varepsilon_2)L-\alpha(1+\gamma)V_0}{\alpha(1-\gamma)}$	$C_g < \varepsilon_1(f+F)$

As demonstrated in Scenarios 1 and 2, platforms may trade user behavioral data even without user authorization. Scenario 3 shows that even with user authorization, platforms may refrain from any data transactions despite being able to trade personal and behavioral data. Thus, user authorization does not invariably promote platform data transactions. In summary, under the user authorization mechanism, platforms' choice of data transaction strategy is closely tied to the transaction price and the value of user behavioral data. As demonstrated by Scenarios 4 and 5, platforms will choose to trade both personal and behavioral data when the transaction price is low and the value of behavioral data exceeds a certain threshold, or when the transaction price is high and the value of behavioral data falls within a specific range. Therefore, only under these conditions can user authorization effectively promote platform data transactions.

As demonstrated in Scenario 2, even without user authorization, the government may still impose strict regulation on the trading of user behavioral data by platforms. As shown in Scenario 4, even with user authorization, the government may still implement loose regulation on the trading of both personal and behavioral user data by platforms. This indicates that user authorization alone is not a sufficient condition for the government to adopt a strict regulatory strategy toward platform data transactions. In summary, under the user authorization mechanism, the government's regulatory strategy for platform data transactions is closely tied to the transaction price, the value of user behavioral data, and regulatory costs. Scenario 5 indicates that a three-party equilibrium where the government imposes strict regulation on platform data transactions under user authorization only emerges when either the transaction price is low, the value of user behavioral data is high, and regulatory costs are low, or the transaction price is high, the value of user behavioral data is moderate, and regulatory costs are low.

5. Evolutionary Simulation Analysis

5.1. Evolutionary Stable Strategies

To further validate the model's correctness, this study employs MATLAB R2023b to simulate the evolutionary process of the three-party strategy, exploring the evolutionary equilibrium of strategies among users, platforms, and government. Based on the actual context of this study, different parameter values are set under conditions $\gamma > 1$ and $0 < \gamma < 1$ for simulation analysis. Specifically, when $\gamma > 1$, the parameter values are as shown in Table 8, and when $0 < \gamma < 1$, the values of each parameter are as shown in Table 9.

Table 8. Parameter settings for scenario $\gamma > 1$.

	Stable Point	α	β	ε_1	ε_2	γ	V_0	ΔV	C	C_g	f	F	L
Scenario 1	(0,1,0)	0.3	0.5	0.5	0.2	2	2	7	3	1	0.6	1.7	2
Scenario 2	(0,1,1)	0.3	0.5	0.5	0.2	2	2	7	3	0.5	3	1.7	2
Scenario 3	(1,0,0)	0.3	0.5	0.5	0.2	2	2	1	3	1	0.6	1.7	2
Scenario 4	(1,1,0)	0.3	0.5	0.5	0.2	2	2	3	3	2	0.6	1.7	2
Scenario 5	(1,1,1)	0.3	0.5	0.5	0.2	2	2	3	3	0.5	0.6	1.7	2

Table 9. Parameter settings for scenario $0 < \gamma < 1$.

	Stable Point	α	β	ε_1	ε_2	γ	V_0	ΔV	C	C_g	f	F	L
Scenario 1	(0,1,0)	0.3	0.5	0.5	0.2	0.6	2	4.5	1	2	1	2	6
Scenario 2	(0,1,1)	0.3	0.5	0.5	0.2	0.6	2	4.5	1	0.5	1	2	6
Scenario 3	(1,0,0)	0.3	0.5	0.5	0.2	0.6	2	0.5	1	2	1	2	6
Scenario 4	(1,1,0)	0.3	0.5	0.5	0.2	0.6	2	8.5	1	2	1	2	6
Scenario 5	(1,1,1)	0.3	0.5	0.5	0.2	0.6	2	8.5	1	1	1	2	6

Based on the above parameter settings, this study simulates the evolutionary trajectory of the evolutionary game system by setting the initial values for each strategy to be 0.1, with increments of 0.4 up to 0.9. The MATLAB R2023b evolutionary paths are shown in Figures 5 and 6. Here, the x, y and z axes represent the probabilities that users (x), platforms (y), and the government (z), respectively, choose the strategy of authorization, transaction, and strict regulation. Each colored line represents the evolutionary path of the system over time, starting from a random initial state. The colors are used solely to distinguish trajectories originating from different initial points and carry no special meaning.

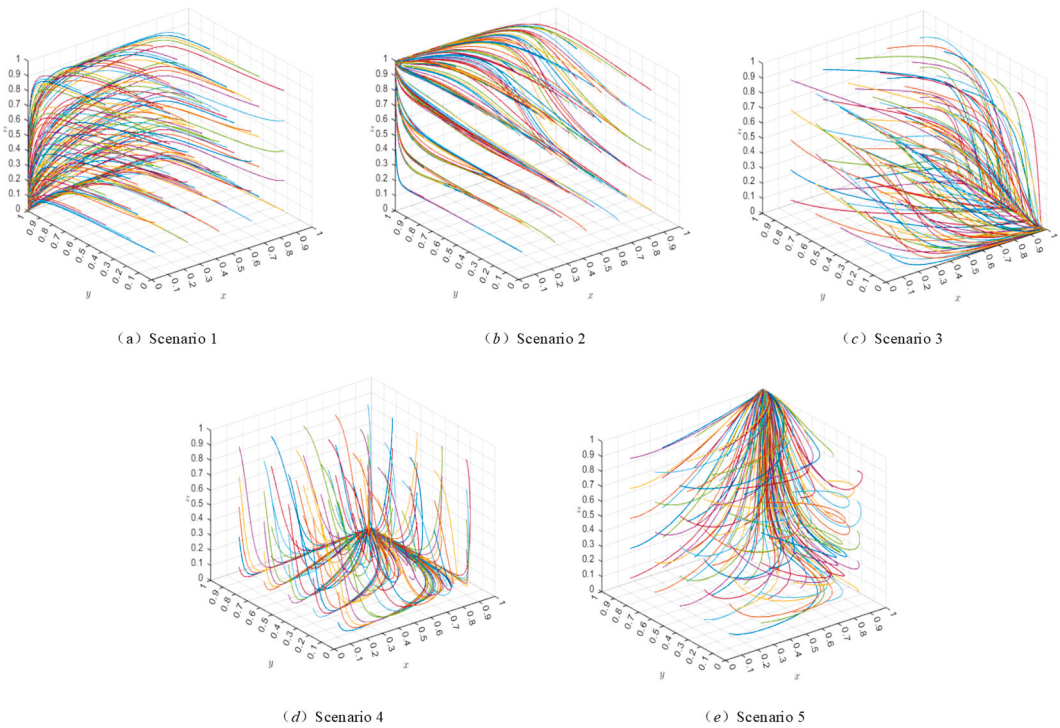


Figure 5. System evolution pathway when $\gamma > 1$.

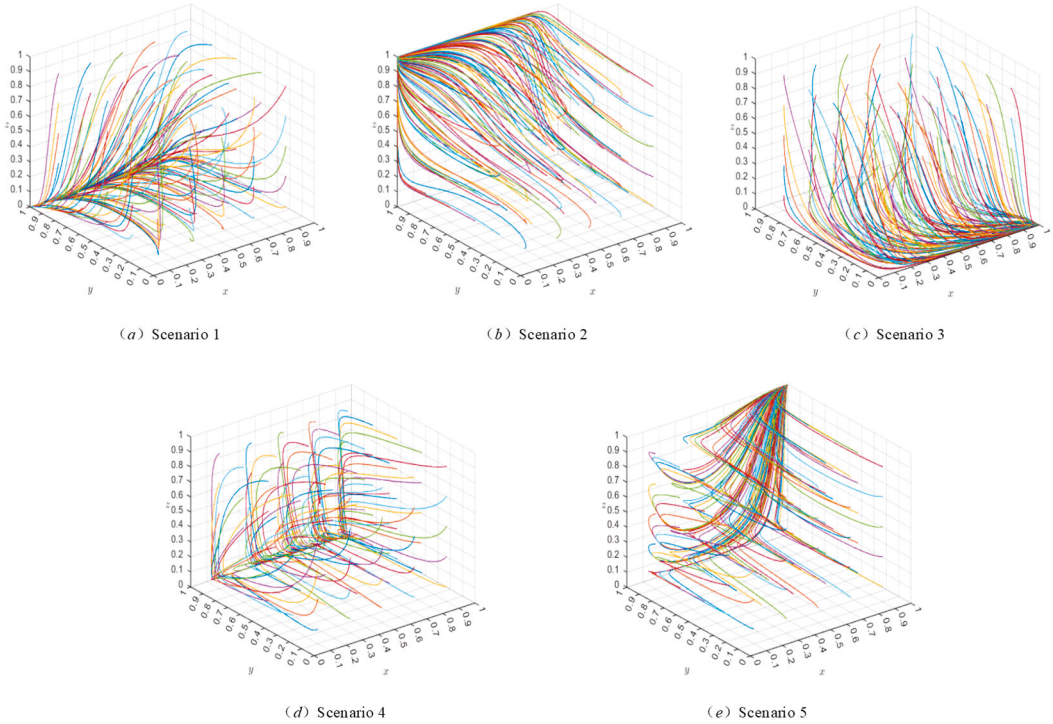


Figure 6. System evolution pathway when $0 < \gamma < 1$.

5.2. Parameter Analysis

Based on the discussion in Section 4.4, simulation analyses were conducted to examine how changes in the value of user behavior data under scenarios of higher or lower data prices affect the evolution of the tripartite strategy equilibrium. The results are as follows.

5.2.1. The Impact of ΔV When $\gamma > 1$

The initial scenario $\Delta V = 1$ was set with a step size of 2, while other parameter values remained consistent with the baseline scenario. The resulting system simulation evolution path is shown in Figure 7.

Figure 7 simulation results indicate that when the value of user behavioral data falls below a certain threshold, the stable strategy combination is (authorization, non-transaction, loose regulation). When the value of user behavioral data exceeds a certain threshold, platforms adopt the transaction strategy. As analyzed in Proposition 5, the probability of government's strict regulation increases continuously until the strict regulation strategy is selected, ultimately reaching the stable strategy (authorization, transaction, strict regulation). As the value of user behavioral data further increases, the stable policy combination evolves to (non-authorization, transaction, strict regulation). This trend indicates that when $\gamma > 1$ and the value of user behavioral data falls within a certain range, platforms' strategy can stabilize on transaction under user authorization; when the value of user behavioral data exceeds this range, platforms' strategy can still stabilize on transaction under non-authorization; however, when the value of user behavioral data falls below this range, platforms' strategy stabilizes on non-transaction. Furthermore, when $\gamma > 1$ and the value of user behavior data falls within a certain range, the government strategy can stabilize at strict regulation when users grant authorization. When the value of user behavior data exceeds this range, the government strategy can still stabilize at strict regulation even when users do not grant authorization.

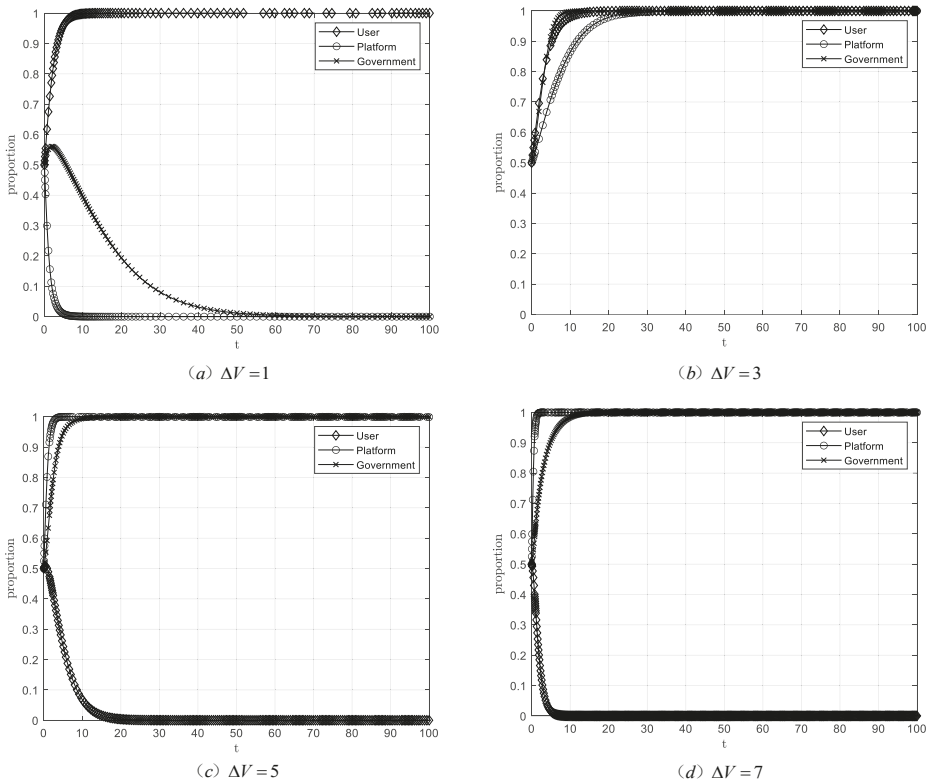


Figure 7. Evolutionary equilibrium when $\gamma > 1$.

5.2.2. The Impact of ΔV When $0 < \gamma < 1$

Modifying the value of user behavior data, under the initial scenario $\Delta V = 0.5$ with a step size of 4 and all other parameters consistent with the baseline scenario, yields the system simulation evolution path shown in Figure 8.

Figure 8 simulation results indicate that when the value of user behavioral data falls below a certain threshold, the stable strategy combination is (authorize, non-transaction, loose regulation). When the value exceeds this threshold, the stable strategy becomes (non-authorization, transaction, strict regulation). As the value of user behavioral data continues to increase, it ultimately reaches a stable point (authorization, transaction, strict regulation). This trend indicates that when $0 < \gamma < 1$ and the value of user behavioral data falls within a certain range, platforms' strategy can stabilize on trading when users do not grant authorization. When the value of user behavioral data exceeds this range, platforms' strategy can still stabilize on transaction when users grant authorization. However, when the value of user behavioral data falls below this range, platforms' strategy stabilizes on non-transaction. When $0 < \gamma < 1$ and the value of user behavioral data falls within a certain range, the government's strategy can stabilize at strict regulation when users do not grant authorization. When the value of user behavioral data exceeds this range, the government strategy can still stabilize at strict regulation when users grant authorization.

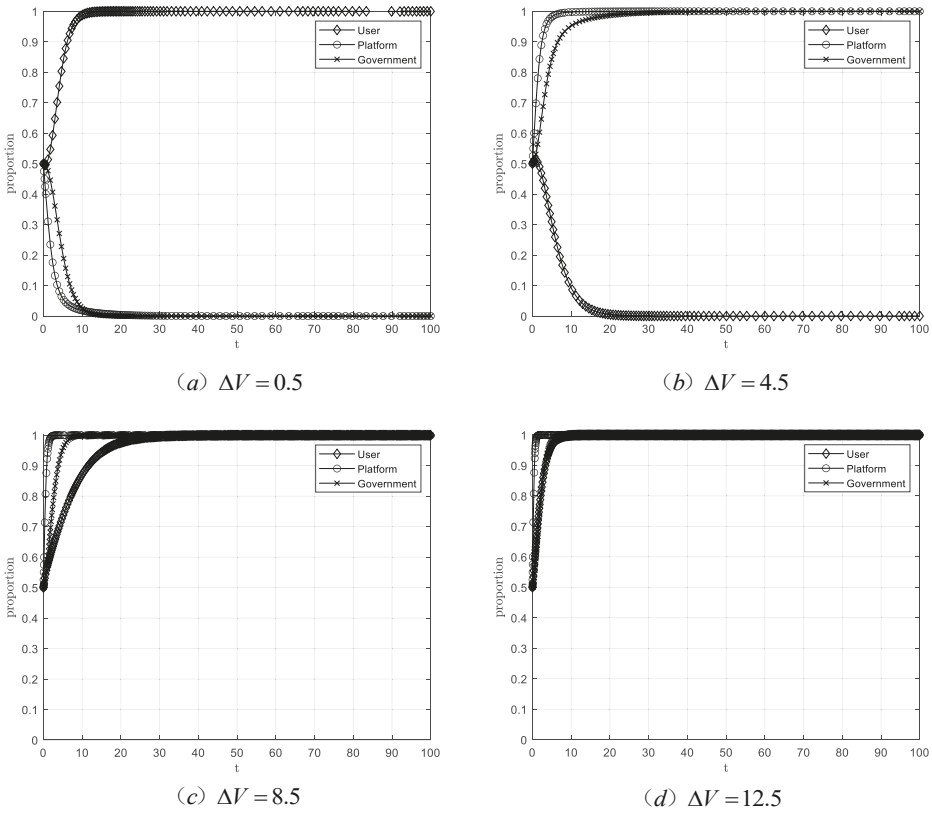


Figure 8. Evolutionary equilibrium when $0 < \gamma < 1$.

6. Conclusions

6.1. Findings

This study builds a tripartite evolutionary game model involving users, platforms, and the government to explore the interactions and evolutionary equilibrium among users' data authorization strategy, platforms' data transaction strategy, and the government's regulation strategy under a user-authorized personal data mechanism. The key findings are as follows:

(1) Under the personal data authorization mechanism, the promotional effect of user authorization on platform data transactions is conditional. The study finds that user authorization does not necessarily enhance the activity of platform data transactions; its effectiveness primarily depends on the data transaction price and the value of user behavioral data. Specifically, user authorization can effectively promote platform data transactions only when the data transaction price is low and the value of user behavioral data exceeds a certain threshold, or when the data transaction price is high and the value of user behavioral data falls within a moderate range. Conversely, user authorization has limited promotional effects on platform data transactions and may even suppress such activities due to data leakage risks. This conclusion indicates that relying solely on user authorization mechanisms cannot fully resolve efficiency issues in data trading markets; synergistic optimization involving both data market supply–demand dynamics and data value assessment is necessary.

(2) Under the personal data authorization mechanism, government regulation of platform data transactions depends on multidimensional factors, with user authorization not constituting a sufficient condition for strict regulatory strategies. Research indicates that whether the government imposes stringent regulation on platform data transactions is influenced not only by user authorization behavior but more critically by the dynamic interplay of regulatory costs, data transaction prices, and the value of user behavioral

data. Specifically, when regulatory costs are low, the government tends to impose strict regulation on platform data transactions under two conditions: when the data price is low and behavioral data value is moderate, or when the data price is high and behavioral data value is high—even without user authorization. Conversely, when user authorization exists, the government leans toward strict regulation under two conditions: when the data price is low and behavioral data value is high, or when the data price is high and behavioral data value is moderate. This indicates that the government's regulatory logic is significantly guided by a "cost–benefit" approach, where user authorization serves only as one of the reference factors in regulatory decision-making, rather than a decisive condition.

6.2. Recommendations

The evolutionary equilibrium of options made by users, platforms, and government policies must foster a virtuous ecosystem characterized by user authorization, platform transactions, and strict government regulation. From the perspective of factors influencing behavioral strategies, the following policy recommendations are proposed:

(1) A standardized data trading market system should be established so that data transaction price γ reflects actual supply–demand dynamics. It is suggested to create a unified data exchange platform with tiered classification standards—categorizing data as foundational, critical, core, etc.—accompanied by differentiated pricing guidelines. Simultaneously, supporting mechanisms for data rights confirmation, registration, and settlement should be refined to effectively evaluate and adjust pricing standards, ensuring alignment with market value.

(2) A scientific data valuation mechanism should be developed to quantify the value of user personal data V_0 and behavioral data ΔV. It is suggested to establish national standards for data valuation, creating multidimensional quantitative models that incorporate factors like data quality and application scenarios. Professional third-party assessment agencies should be cultivated to provide objective value references for data transactions. Simultaneously, platform enterprises should be encouraged to implement dynamic pricing based on assessment results and enhance data transparency by clearly disclosing data flow and usage to users.

(3) Data breach prevention and control should be strengthened by optimizing penalty f and regulatory costs C_g. Tiered fines linked to data value and harm severity should be implemented to impose strict penalties for high-value data breaches or major violations. Blacklist and credit punishment mechanisms should be refined to conduct regular security inspections, and data trading parties should be urged to implement protective measures like encryption and access controls to enhance overall security standards.

6.3. Limitations and Future Research

This study presents a tripartite evolutionary game model focusing on the interactions between users, platforms and the government under a user data authorization mechanism, and examines their strategies and evolutionary equilibrium. However, the model is primarily based on the characteristics of China's data trading market and involves certain contextual simplifications. The universality of its conclusions across different institutional environments therefore warrants careful scrutiny. Specifically, the model framework may have limitations when applied to different data governance systems, such as the General Data Protection Regulation (GDPR). Additionally, elements unique to digital markets, including multiple user groups and network effects, remain outside the analytical scope of the current model. Therefore, future research could test the model within broader institutional and market contexts. Particular emphasis should be placed on validating and expanding

the theoretical findings of this study in real-world scenarios involving cross-border data flows and the interaction of multiple governance frameworks.

Author Contributions: Conceptualization and methodology, Z.L. and J.L.; writing, Z.L. and J.L.; supervision, Z.L.; numerical simulation, J.L., M.Q. and Z.C.; investigation and visualization, J.L., M.Q. and Z.C.; Project administration and writing—review and editing, Z.L. and J.L.; Supervision and writing—review and editing, Z.L. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Social Science Foundation of China (grant number 25BGL127).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors would like to thank the editors and the anonymous reviewers for their insightful suggestions to improve the quality of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, Q.; Serfes, K. Customer Information Sharing among Rival Firms. *Eur. Econ. Rev.* **2006**, *50*, 1571–1600. [CrossRef]
2. Jones, C.I.; Tonetti, C. Nonrivalry and the Economics of Data. *Am. Econ. Rev.* **2020**, *110*, 2819–2858. [CrossRef]
3. Acquisti, A.; Taylor, C.; Wagman, L. The Economics of Privacy. *J. Econ. Lit.* **2016**, *54*, 442–492. [CrossRef]
4. Veldkamp, L.; Chung, C. Data and the Aggregate Economy. *J. Econ. Lit.* **2024**, *62*, 458–484. [CrossRef]
5. Gaessler, F.; Wagner, S. Patents, Data Exclusivity, and the Development of New Drugs. *Rev. Econ. Stat.* **2022**, *104*, 571–586. [CrossRef]
6. Varian, H.R. *Artificial Intelligence, Economics, and Industrial Organization*; National Bureau of Economic Research: Cambridge MA, USA, 2018.
7. Tomak, K.; Keskin, T. Exploring the Trade-off Between Immediate Gratification and Delayed Network Externalities in the Consumption of Information Goods. *Eur. J. Oper. Res.* **2008**, *187*, 887–902. [CrossRef]
8. Abowd, J.M.; Schmutte, I.M. An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices. *Am. Econ. Rev.* **2019**, *109*, 171–202. [CrossRef]
9. Dosis, A.; Sand-Zantman, W. The Ownership of Data. *J. Law. Econ. Organ.* **2023**, *39*, 615–641. [CrossRef]
10. Miller, A.R.; Tucker, C. Privacy Protection, Personalized Medicine, and Genetic Testing. *Manag. Sci.* **2018**, *64*, 4648–4668. [CrossRef]
11. Besbes, O.; Mouchtaki, O. How Big Should Your Data Really Be? Data-Driven Newsvendor: Learning One Sample at A Time. *Manag. Sci.* **2023**, *69*, 5848–5865. [CrossRef]
12. Bar-Gill, S.; Brynjolfsson, E.; Hak, N. Helping Small Businesses Become More Data-Driven: A Field Experiment on EBay. *Manag. Sci.* **2024**, *70*, 7345–7372. [CrossRef]
13. Gupta, M.; George, J.F. Toward the Development of A Big Data Analytics Capability. *Inf. Manag.* **2016**, *53*, 1049–1064. [CrossRef]
14. Zhang, R.; Zhu, H.; Chang, Q.; Mao, Q. A Comprehensive Review of Digital Twins Technology in Agriculture. *Agriculture* **2025**, *15*, 903. [CrossRef]
15. Taha, M.F.; Mao, H.; Zhang, Z.; Elmasry, G.; Awad, M.A.; Abdalla, A.; Mousa, S.; Elwakeel, A.E.; Elsherbiny, O. Emerging Technologies for Precision Crop Management towards Agriculture 5.0: A Comprehensive Overview. *Agriculture* **2025**, *15*, 582. [CrossRef]
16. Cong, L.W.; Xie, D.; Zhang, L. Knowledge Accumulation, Privacy, and Growth in A Data Economy. *Manag. Sci.* **2021**, *67*, 6480–6492. [CrossRef]
17. Quach, S.; Thaichon, P.; Martin, K.D.; Weaven, S.; Palmatier, R.W. Digital Technologies: Tensions in Privacy and Data. *J. Acad. Market. Sci.* **2022**, *50*, 1299–1323. [CrossRef]
18. Choe, C.; Cong, J.; Wang, C. Softening Competition Through Unilateral Sharing of Customer Data. *Manag. Sci.* **2024**, *70*, 526–543. [CrossRef]
19. Chen, N.; Cire, A.A.; Hu, M.; Lagzi, S. Model-Free Assortment Pricing with Transaction Data. *Manag. Sci.* **2023**, *69*, 5830–5847. [CrossRef]

20. Lei, Y.; Miao, S.; Momot, R. Privacy-Preserving Personalized Revenue Management. *Manag. Sci.* **2024**, *70*, 4875–4892. [CrossRef]
21. Allouah, A.; Bahamou, A.; Besbes, O. Optimal Pricing with A Single Point. *Manag. Sci.* **2023**, *69*, 5866–5882. [CrossRef]
22. Beraja, M.; Yang, D.Y.; Yuchtman, N. Data-Intensive Innovation and the State: Evidence from AI Firms in China. *Rev. Econ. Stud.* **2023**, *90*, 1701–1723. [CrossRef]
23. Akcigit, U.; Celik, M.A.; Greenwood, J. Buy, Keep, or Sell: Economic Growth and the Market for Ideas. *Econometrica* **2016**, *84*, 943–984. [CrossRef]
24. Lee, Y.S.; Weber, R.A. Revealed Privacy Preferences: Are Privacy Choices Rational? *Manag. Sci.* **2025**, *71*, 2657–2677. [CrossRef]
25. Pei, J. A Survey on Data Pricing: From Economics to Data Science. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 4586–4608. [CrossRef]
26. Lerner, J.; Pathak, P.A.; Tirole, J. The Dynamics of Open-Source Contributors. *Am. Econ. Rev.* **2006**, *96*, 114–118. [CrossRef]
27. Wu, S.; Banker, R. Best pricing strategy for information services. *J. Assoc. Inf. Syst.* **2010**, *11*, 339–366. [CrossRef]
28. Bergemann, D.; Bonatti, A. Selling Cookies. *Am. Econ. J. Microecon.* **2015**, *7*, 259–294. [CrossRef]
29. Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent Trends in Deep Learning Based Natural Language Processing. *IEEE Comput. Intell. Mag.* **2018**, *13*, 55–75. [CrossRef]
30. Bauer-Hänsel, I.; Liu, Q.; Tessone, C.J.; Schwabe, G. Designing A Blockchain-Based Data Market and Pricing Data to Optimize Data Trading and Welfare. *Int. J. Electron. Commer.* **2024**, *28*, 3–30. [CrossRef]
31. Abbas, A.E.; Agahari, W.; van de Ven, M.; Zuiderwijk, A.; de Reuver, M. Business Data Sharing through Data Marketplaces: A Systematic Literature Review. *J. Theor. Appl. Electron. Commer. Res.* **2021**, *16*, 3321–3339. [CrossRef]
32. Koutroumpis, P.; Leiponen, A.; Thomas, L.D.W. Markets for Data. *Ind. Corp. Change* **2020**, *29*, 645–660. [CrossRef]
33. Zhang, X.; Zha, X.; Zhang, H.; Dan, B. Information Sharing in A Cross-Border E-Commerce Supply Chain Under Tax Uncertainty. *Int. J. Electron. Commer.* **2022**, *26*, 123–146. [CrossRef]
34. Johnson, J.P.; Jungbauer, T.; Preuss, M. Online Advertising, Data Sharing, and Consumer Control. *Manag. Sci.* **2024**, *70*, 7984–8002. [CrossRef]
35. Bergemann, D.; Bonatti, A.; Gan, T. The Economics of Social Data. *RAND J. Econ.* **2022**, *53*, 263–296. [CrossRef]
36. Jorzik, N.; Kirchhof, P.J.; Mueller-Langer, F. Industrial Data Sharing and Data Readiness: A Law and Economics Perspective. *Eur. J. Law. Econ.* **2024**, *57*, 181–205. [CrossRef]
37. Agarwal, S.; Ghosh, P.; Ruan, T.; Zhang, Y. Transient Customer Response to Data Breaches of Their Information. *Manag. Sci.* **2024**, *70*, 4105–4114. [CrossRef]
38. Alonso, R.; Câmara, O. Organizing Data Analytics. *Manag. Sci.* **2024**, *70*, 3123–3143. [CrossRef]
39. Urbonavicius, S. Relative Power of Online Buyers in Regard to A Store: How It Encourages Them to Disclose Their Personal Data? *J. Retail. Consum. Serv.* **2023**, *75*, 103510. [CrossRef]
40. Luo, Y.; Zhou, L.; Huang, J.; Wang, X.; Sun, R.; Zhu, G. Platform Perspective Versus User Perspective: The Role of Expression Perspective in Privacy Disclosure. *J. Retail. Consum. Serv.* **2023**, *73*, 103372. [CrossRef]
41. Loebbecke, C.; van Fenema, P.C.; Powell, P. Managing Inter-Organizational Knowledge Sharing. *J. Strateg. Inf. Syst.* **2016**, *25*, 4–14. [CrossRef]
42. Notheisen, B.; Cholewa, J.B.; Shanmugam, A.P. Trading Real-World Assets on Blockchain: An Application of Trust-Free TransAction Systems in the Market for Lemons. *Bus. Inform. Syst. Eng.* **2017**, *59*, 425–440. [CrossRef]
43. Pavlou, P.A.; Gefen, D. Building Effective Online Marketplaces with Institution-Based Trust. *Inform. Syst. Res.* **2004**, *15*, 37–59. [CrossRef]
44. Wang, E.S.-T. Role of Privacy Legislations and Online Business Brand Image in Consumer Perceptions of Online Privacy Risk. *J. Theor. Appl. Electron. Commer. Res.* **2019**, *14*, 59–69. [CrossRef]
45. Sun, T.; Yuan, Z.; Li, C.; Zhang, K.; Xu, J. The Value of Personal Data in Internet Commerce: A High-Stakes Field Experiment on Data Regulation Policy. *Manag. Sci.* **2024**, *70*, 2645–2660. [CrossRef]
46. Kang, J.; Lan, J.; Huang, S.; Chen, L. Effects of Privacy Regulatory Protection on Users' Data Sharing in Mobile Apps. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 153. [CrossRef]
47. Marikyan, D.; Papagiannidis, S.; Rana, O.F.; Ranjan, R. General Data Protection Regulation: A study On Attitude and Emotional Empowerment. *Behav. Inform. Technol.* **2024**, *43*, 3561–3577. [CrossRef]
48. Mahajan, S. Data Marketplaces: A Solution for Personal Data Control and Ownership? *Sustainability* **2020**, *14*, 16884. [CrossRef]
49. Li, C.; Li, H.; Tao, C. Evolutionary Game of Platform Enterprises, Government and Consumers in the Context of Digital Economy. *J. Bus. Res.* **2023**, *167*, 113858. [CrossRef]
50. D'Hauwers, R.; Vandercruysse, L. Competitive Advantage and Personal Data Ecosystems: A Typology of Personal Data Control Constellations. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 8. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Is a Chatbot More Effective? Investigating the Effect of Service Recovery Agents and Consumer Loss on Consumer Forgiveness

Liu Fan [1], Shanshan Li [1], Can Wang [1] and Xiaoping Zhang [2,*]

[1] Management College, Ocean University of China, 238 Songling Road, Laoshan District, Qingdao 266100, China; imfanliu@163.com (L.F.); lshanshan0808@163.com (S.L.)
[2] School of Economics, Wuhan University of Technology, 122 Luoshi Road, Hongshan District, Wuhan 430070, China
* Correspondence: xiaoping_sdust@163.com

Abstract

As chatbots are increasingly deployed to address service failures, understanding their role in facilitating consumer forgiveness has become essential. Several studies have compared consumers' reactions to service recovery efforts conducted by a human versus a chatbot. Through three scenario-based experiments (total N = 1875) with Chinese participants, our study examines the interaction between service recovery agents (chatbot vs. human), types of consumer loss (utilitarian vs. symbolic), and service failure severity (low vs. high) in influencing consumer forgiveness. The results reveal that in cases of symbolic loss, consumers perceive humans—rather than chatbots—as more capable of providing emotional support during service recovery, thus promoting forgiveness more effectively. However, this discrepancy diminishes in the case of utilitarian loss. Our findings further suggest that the combined effect of service recovery agents and consumer loss on forgiveness is moderated by service failure severity. In the case of low-severity failures, recovery services provided by humans (vs. chatbots) are more effective in fostering forgiveness for consumers experiencing symbolic losses. However, for high-severity failures, regardless of the type of loss, consumers exhibit a higher level of forgiveness toward recovery services provided by humans. This research offers the following practical implications for managers dealing with service failures: strategic escalation to human agents is recommended for symbolic losses or high-severity failures, but chatbots represent a cost-efficient solution for utilitarian losses in low-severity scenarios.

Keywords: service recovery agent; consumer loss; perceived responsiveness; perceived emotional support; consumer forgiveness

1. Introduction

Service failure occurs when a firm's or brand's actual performance falls short of consumers' expected standards [1]. This failure inevitably leads to consumer loss, which may include financial damages, wasted time, and emotional exhaustion [2–4]. In response, firms implement strategies to redress these losses, a process known as service recovery [5]. Effective service recovery strategies, such as apologies, compensation, explanations, and humor-based communication, can mitigate the negative effects of service failure, which is crucial for enhancing consumer satisfaction and loyalty [6–8]. Notably, a high level of consumer forgiveness is a key indicator of successful service recovery [9]. Consumer forgiveness refers to the act of consumers releasing grudges or anger and adopting a more favorable attitude toward the offending company [10].

In traditional service recovery contexts, human agents serve as recovery service providers to manage service failures [11]. By offering emotional support and personalized solutions, they help alleviate consumer dissatisfaction during the service recovery process. Recently, AI-enabled chatbots have been increasingly deployed to complement human-based consumer services [12–14]. Compared to human-based services, chatbot-based solutions offer distinct advantages, including lower costs and 24/7 availability [15]. More importantly, with their growing ability to understand consumer needs and emotions, chatbots are becoming capable of serving as recovery agents and engaging in the service recovery process to address service failures. This study aims to investigate whether chatbots can effectively handle service failures in a manner similar to humans and contribute to increased consumer forgiveness.

Numerous studies have shown that individuals' attitudes toward services provided by humans versus AI depend on the nature of the task [16–18]. For example, consumers generally perceive AI agents as more capable of assessing objective and utilitarian attributes [17,19]. Extending these findings to our research context, we predict that consumers' responses—specifically consumer forgiveness—to services provided by a chatbot versus a human will depend on the type of consumer loss resulting from the service failure. We propose that service failure can lead to two distinct types of consumer losses: utilitarian and symbolic losses [20]. Utilitarian losses typically involve economic resources, while symbolic losses pertain to psychological and social resources. Existing literature primarily explores consumers' reactions to chatbot service failures [21–24], but largely overlooks how consumers respond to chatbot involvement in service recovery following service failures. Although recent studies examine consumers' perceptions and intentions regarding chatbots' role in service recovery [22,25], there is limited research on how consumers' reactions to service recovery vary when administered by a human versus a chatbot across different types of consumer losses. To address this gap, the first objective of this study is to investigate how the impact of service recovery agents on consumer forgiveness is influenced by the type of consumer loss.

Additionally, service failure severity is a frequently discussed concept in the literature on service failures, as it significantly influences consumers' attitudes toward recovery efforts [26,27]. Service failure severity can shape consumers' perceptions and behavioral responses within service failure contexts [28,29]. Therefore, the second research objective of this study is to examine how the interaction between service recovery agents and consumer loss is influenced by the severity of the service failure.

Several key contributions emerge from this research. First, this study represents an initial effort to examine the interplay between service recovery agents and types of consumer loss in shaping forgiveness following a service failure, thereby enriching the existing marketing literature on service failure and recovery. Second, it reveals two mediating pathways—perceived responsiveness and perceived emotional support—that explain the underlying mechanisms through which recovery agents and loss types influence forgiveness. Finally, we identify a critical boundary condition for this interaction by demonstrating that service failure severity moderates the effects of the recovery agent and the type of loss.

2. Literature Review and Research Model

2.1. Service Failure and Consumer Forgiveness

Service failure is defined as "any instance of actual or perceived mishap, error, or problem that occurs during the experience phase" [30]. It typically arises when the service received does not meet a consumer's expectations. For example, in the hospitality and tourism sectors, consumers may encounter issues such as hotel noise, stains on bathtubs

https://doi.org/10.3390/jtaer21010035

or shower curtains, and poor attitudes from service employees [31]. In the foodservice industry, service failures often result from delays in food delivery, foreign objects found in a customer's dish, or rude or indifferent behavior from servers [32]. Service failures generally lead to negative consequences, including consumer dissatisfaction, decreased consumer confidence, and negative word-of-mouth (WOM) [33].

Service failure inevitably leads to customer losses. Grönroos (1984) argued that service failure can, on the one hand, cause consumers to incur economic losses, and on the other hand, result in emotional losses [34]. Building on this, Du and Fan (2007) proposed two types of consumer loss: utilitarian loss and symbolic loss [20]. The distinction between these two types of losses is summarized in Table 1. Utilitarian loss refers to tangible flaws or functional deficiencies that diminish the practical value of a product. It typically involves economic resources such as money, goods, and time, focusing on the losses consumers experience in terms of actual outcomes [20]. Symbolic loss, by contrast, pertains to psychological or social deficiencies that may damage a consumer's identity, dignity, or relational value. It generally involves psychological and social resources such as status, respect, and empathy. Thus, symbolic loss often emphasizes the consumer's perception of the relationship between themselves and the service provider following the service failure [20]. Consumers who experience utilitarian losses are typically more concerned with how service providers address their practical problems [4]. In contrast, symbolic losses often involve emotional discomfort, such as feelings of disrespect, neglect, and dissatisfaction. Consumers facing symbolic losses may perceive harm to their image, social status, and personal value [35].

Table 1. Difference between Utilitarian Loss and Symbolic Loss.

Core Feature	Utilitarian Loss	Symbolic Loss
Definition	Obvious and significant flaws or functional deficiencies that reduce the practical value of a service	Psychological or social deficiencies that may harm the consumer's identity, dignity, or relational value
Focus	Outcome-relevant threats (e.g., practical, transactional, etc.)	Identity-relevant threats (e.g., relational, self-relate, etc.)
Features	Tangible, measurable, functional, resource-based	Intangible, subjective, relational, norm-based

When consumers experience service losses, merchants typically implement a range of service recovery strategies. Consumer forgiveness is a crucial outcome of effective service recovery and is defined as the consumer's willingness to release resentment toward a firm and adopt a constructive, trusting stance following a service failure [9,10]. Existing literature conceptualizes forgiveness as a multifaceted process influenced by several factors. For example, perceived justice, which encompasses distributive, procedural, and interactional dimensions, serves as a key driver, as consumers are more likely to forgive when they perceive recovery efforts as fair [10]. Joireman et al. (2016) emphasized that the severity of service failure amplifies negative emotions, thus reducing the likelihood of forgiveness by lowering consumers' tolerance for errors made by firms [36]. While previous studies have enhanced our understanding of how to promote consumer forgiveness, these findings may not be directly applicable to situations in which chatbots serve as the service recovery agent. This study focuses on the combined role of consumer loss type (utilitarian vs. symbolic) and service failure severity in influencing preferences for recovery agents. However, it remains unclear whether chatbots' ability to foster consumer forgiveness varies across different types of losses and service failure severity.

2.2. Responses to Decisions and Services Provided by a Human Versus AI

Numerous studies have examined consumers' reactions to services provided by AI versus humans. Existing literature suggests that, although AI often performs better than humans, consumers tend to prefer human decisions over those made by AI, a phenomenon known as algorithm aversion [19,37,38]. However, researchers have found that individuals' attitudes toward AI and human decision-making are influenced by task characteristics and the favorability of the decision [39–41]. For example, people tend to prefer algorithms over human advice when dealing with objective tasks, a phenomenon referred to as algorithm appreciation [19,42,43].

In the context of service failures, most studies have focused on consumers' attitudes toward service failures caused by AI-enabled agents. For example, research has shown that consumers report lower satisfaction and purchase intentions when a virtual agent (vs. a human agent) causes a service failure [44]. Zhang et al. (2025) found that a consumer's intention to continue using chatbots is influenced by the chatbot's emotional expression during service failures [8]. Recently, scholars have begun to explore how chatbots, as service recovery agents, impact consumers' perceptions, intentions, and behaviors. These studies suggest that factors such as chatbots' language style, anthropomorphism, empathy, and emoji types influence consumer service recovery satisfaction and forgiveness [45–47]. For instance, Cao and Yu (2026) found that chatbots using facial emojis lead to higher consumer satisfaction during service recovery than those using non-facial emojis [48]. Similarly, Shams et al. (2024) discovered that chatbots using informal and human-like language increase consumer satisfaction with service recovery efforts [45]. These studies also highlight the underlying mechanisms by which chatbots promote consumer forgiveness, including perceived empathy, perceived ambiguity, trustworthiness, and perceived governance [46,47,49]. However, the existing literature has paid less attention to how different types of consumer loss influence consumer forgiveness. Specifically, the mediating roles of perceived emotional support and perceived responsiveness in consumer forgiveness, particularly when comparing chatbots to humans, depending on the type of loss (utilitarian vs. symbolic), have not yet been empirically verified.

2.3. Task-Technology Fit Model: Interaction Between Consumer Loss and Service Recovery Agent

Consumer forgiveness is achieved through both emotional and cognitive pathways [50]. Drawing on existing literature, this study hypothesizes two mediating pathways—perceived emotional support (i.e., emotional pathway) and perceived responsiveness (i.e., cognitive pathway). Perceived emotional support is defined as the provision of empathetic and reassuring expressions that help alleviate individuals' emotional distress [51]. Perceived responsiveness, on the other hand, refers to the speed and accuracy with which a service provider delivers prompt, efficient, and relevant information in response to consumer needs [52].

We next draw on task-technology fit theory to explore how the interaction between recovery agents and loss type influences consumer forgiveness. Task-technology fit theory posits that the alignment between task requirements and technology attributes can significantly influence technology utilization and user performance. Task-technology fit refers to the degree to which technology aids an individual in performing a task [53]. This fit is determined by two key components: task attributes and technology attributes. When a technology's attributes closely match the demands of a task, task-technology fit is high, leading to positive outcomes [53]. Conversely, even if a technology is advanced, users are unlikely to adopt it if it does not meet their task requirements [54].

In our context, the technology can be understood as either a human agent or a chatbot, while the task is to address consumer losses and facilitate forgiveness. Consumers may

experience utilitarian and/or symbolic losses when encountering service failures. When consumers face symbolic losses, they are more likely to seek emotional value during the service recovery process [55]. They typically desire understanding and empathy from service providers, which helps alleviate the negative emotions triggered by service failures. In such cases, the service recovery provider must address the emotional needs of consumers experiencing symbolic losses. Human service providers, with their inherent capacity for emotional intelligence [55–57], are well-suited to meet these task demands. In contrast, chatbots, which are often perceived as lacking minds and agency, are viewed merely as tools serving humans [58–60]. As a result, they are commonly perceived as lacking empathy and struggling to perform emotion-oriented tasks leading to a lower task-technology fit and diminished effectiveness in fostering forgiveness [17,61,62]. Therefore, in the case of symbolic losses, humans are likely to outperform chatbots in promoting consumer forgiveness by providing more emotional support.

When consumers experience service failures, they generally expect service providers to respond promptly and effectively [63]. In particular, for utilitarian losses involving tangible or functional deficits, consumers place greater emphasis on service providers' adaptability and the provision of practical solutions, as they have higher expectations for the speed of service recovery. In such situations, consumers seek immediate remedies from recovery agents to minimize their losses [64,65]. Today, chatbots have been trained to recognize consumers' essential needs and respond to complaints with similar speed and efficiency as human agents [66–68]. Therefore, in the case of utilitarian losses, both humans and chatbots can meet the demands for quick, tangible solutions and perform effectively in enhancing consumers' perceived responsiveness. As a result, there may be little difference in consumer forgiveness when either chatbots or humans handle utilitarian losses.

Our theoretical model is closely aligned with perceived justice theory, a cornerstone of service recovery research [1,6]. Perceived emotional support reflects interactional justice, which pertains to the fairness of interpersonal treatment during recovery, such as respect, empathy, and sincerity [10]. For consumers experiencing symbolic loss, interactional justice is particularly important, as emotional support addresses the core injustice of being devalued. Perceived responsiveness corresponds with procedural justice, which concerns the fairness of recovery processes, including speed, efficiency, and transparency [69]. In the case of utilitarian loss, procedural justice plays a crucial role, as consumers prioritize the timely resolution of practical issues. We thus hypothesize:

H1a: *In symbolic loss contexts, recovery services provided by humans are more effective in promoting consumer forgiveness than those provided by chatbots.*

H1b: *In utilitarian loss contexts, the effect of the service recovery agent on consumer forgiveness is mitigated.*

H2a: *The effect proposed in H1a is mediated by perceived emotional support. Specifically, in symbolic loss contexts, humans are perceived as more capable of providing emotional support than chatbots, thereby promoting consumer forgiveness.*

H2b: *The effect proposed in H1b is mediated by perceived responsiveness. Specifically, in utilitarian loss contexts, both humans and chatbots are perceived to have high levels of responsiveness, thereby promoting consumer forgiveness.*

2.4. Moderating Effect of Service Failure Severity

Service failure severity refers to the extent of consumer losses [28,70], and it can significantly influence the effectiveness of service recovery [71,72]. The severity of a failure shapes consumers' tolerance, satisfaction, and behavioral responses [73].

Cognitive appraisal theory further explains why service failure severity moderates our core effects [36,69]. Consumers appraise service failures along three key dimensions: controllability (whether the failure was avoidable), stability (how likely it is to recur), and intentionality (whether the failure was deliberate). For high-severity failures, consumers are more likely to appraise the incident as less controllable, more stable, or potentially intentional, which intensifies their feelings of injustice and the need for meaningful redress. In such situations, human agents are better equipped to address these appraisals. In contrast, for low-severity failures, appraisals tend to be less intense (e.g., perceived as accidental or one-time occurrences), so the primary driver of agent preferences remains the core distinction between loss types and their corresponding justice needs.

In cases of severe service failures, consumers experience relatively greater utilitarian and symbolic losses, which intensifies their negative reactions. When service failures are significant, consumers are less likely to forgive service providers [36]. Severe service failures increase consumers' focus on how firms resolve the issue, prompting them to invest more emotional energy in the recovery process [74]. Specifically, when the severity of a service failure is high, consumers expect service providers to address their complaints with sincerity. Furthermore, consumers who experience serious service failures care not only about the timeliness of solutions but also about whether service providers can offer emotional support. As noted earlier, chatbots are perceived as lacking emotional understanding [75], which may hinder forgiveness among consumers experiencing severe service failures. In contrast, humans can establish emotional connections with these consumers, helping to restore their confidence in both the functional and emotional capabilities of service providers. We thus hypothesize:

H3a: *Under the high-severity condition, consumers prefer humans to provide recovery services regardless of the type of consumer loss.*

H3b: *Under the low-severity condition, the interaction effect predicted in H1 remains significant.*

Figure 1 shows the theoretical model.

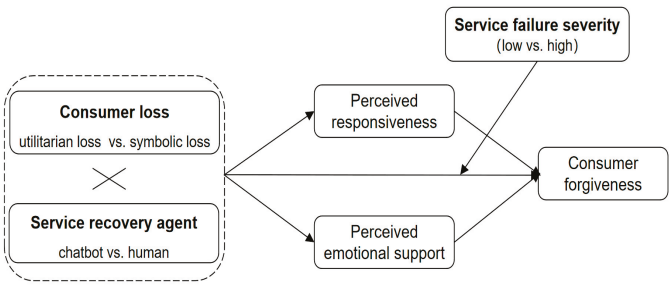


Figure 1. Research model.

3. Study 1

3.1. Design and Participants

Study 1 aimed to examine the interaction between consumer loss and service recovery agent on consumer forgiveness (H1). A 2 (consumer loss: symbolic loss vs. utilitarian loss) × 2 (recovery agent: chatbot vs. human) between-subjects design was employed. A total of 270 participants were recruited through Wenjuanxing and were offered a reward

of 2 yuan. After excluding 30 invalid responses (e.g., duplicate answers, failed attention checks), 240 valid responses were retained (M_{age} = 24.89; 59.7% female).

3.2. Procedure

Consumers' current emotions can influence their forgiveness during service failures [25]. To mitigate the potential impact of participants' current emotions on the results, participants were first asked to report their current emotional state [76]. They also provided feedback on their previous experiences interacting with chatbots. Subsequently, participants were randomly assigned to one of two conditions (utilitarian loss vs. symbolic loss) and read a service failure scenario set in the hotel check-in context. In the symbolic loss condition, participants were informed that they encountered poor service attitudes, while in the utilitarian loss condition, they were told about luggage damage during check-in at Hotel YOUJIA. The detailed experimental materials are provided in Appendix A.1. Participants were then asked to rate their perceived losses [77]. Following this, they were told that they had made a complaint, which was addressed either by a human agent (in the human condition) or by a chatbot (in the chatbot condition). Participants were presented with a screenshot of the dialogue corresponding to their interaction with the human agent or chatbot (see Figures A1 and A2 in Appendix A.1). Finally, they were asked to rate their perceived forgiveness on a 7-point scale. Demographic information was also collected, with all items listed in Appendix B. The reliability and validity of the measurement scales were assessed through confirmatory factor analysis (see Table A2 in Appendix C.1).

3.3. Results

Manipulation Checks. First, we assessed the realism of the experimental scenarios. Participants in both groups, who were exposed to different types of consumer losses, rated the experimental scenario as highly realistic ($M_{symbolic}$ = 5.81, t(119) = 16.32, $p < 0.001$; $M_{utilitarian}$ = 5.06, t(119) = 10.86, $p < 0.001$). Second, the service recovery agent was more frequently perceived as a chatbot in the chatbot group and more frequently perceived as a human in the human group ($M_{chatbot}$ = 1.86, SD = 1.03; M_{human} = 6.09, SD = 1.16; t(238) = −26.73, $p < 0.001$), indicating successful manipulation of the service recovery agent. Regarding consumer loss, participants reported greater utilitarian loss than symbolic loss in the utilitarian loss condition ($M_{symbolic}$ = 5.23, SD = 1.08; $M_{utilitarian}$ = 4.50, SD = 1.41; t(238) = −4.53, $p < 0.001$), whereas they reported greater symbolic loss than utilitarian loss in the symbolic loss condition ($M_{symbolic}$ = 4.51, SD = 1.05; $M_{utilitarian}$ = 5.08, SD = 1.08; t(238) = 4.13, $p < 0.001$). These results confirm successful manipulation of the consumer loss variable. Additionally, participants across the four groups showed no significant differences in their current emotions (F(1, 236) = 3.06, $p > 0.05$).

Interaction between Service Recovery Agent and Consumer Loss. An ANOVA revealed a significant main effect of consumer loss ($M_{utilitarian}$ = 4.36, SD = 1.25; $M_{symbolic}$ = 3.88, SD = 1.25; F(1, 236) = 8.69, $p < 0.05$, $\eta^2 = 0.04$), while the service recovery agent did not have a significant main effect (M_{human} = 4.31, SD = 1.20; $M_{chatbot}$ = 3.92, SD = 1.32; F(1, 236) = 5.99, $p > 0.05$, $\eta^2 = 0.02$) on consumer forgiveness. The 2 × 2 ANOVA revealed a significant interaction between service recovery agent and consumer loss (F(1, 236) = 4.31, $p < 0.05$, $\eta^2 = 0.02$; see Figure 2). In the case of utilitarian loss, the type of service recovery agent had no significant effect on consumer forgiveness (M_{human} = 4.39, SD = 1.03; $M_{chatbot}$ = 4.32, SD = 1.14; F(1, 236) = 0.08, $p > 0.05$). However, in the case of symbolic loss, humans outperformed chatbots in promoting consumer forgiveness (M_{human} = 4.24, SD = 1.09; $M_{chatbot}$ = 3.53, SD = 1.21; F(1, 236) = 10.46, $p < 0.001$). Therefore, H1a and H1b were supported.

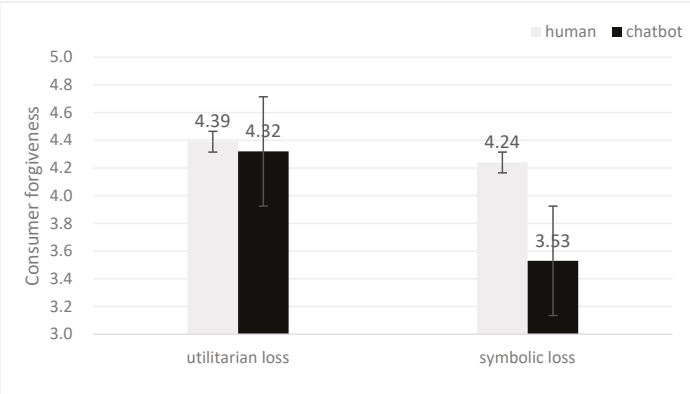


Figure 2. Interaction between service recovery agent and consumer loss on consumer forgiveness.

4. Study 2

4.1. Design and Participants

Study 2 aimed to examine the psychological mechanisms that differentiate consumers' responses to human agents and chatbots in the context of utilitarian or symbolic loss (H2). A 2 (consumer loss: symbolic loss vs. utilitarian loss) × 2 (recovery agent: chatbot vs. human) between-subjects design was employed. A total of 566 participants were recruited through Wenjuanxing and received a reward of 2 yuan. After excluding 46 invalid responses (e.g., duplicate answers, failed attention checks, etc.), 520 valid samples were retained (M_{age} = 23.72 years; 65% female).

4.2. Procedure

Following the procedure used in Study 1, participants first reported their current emotions and their experiences interacting with chatbots. Next, participants were randomly assigned to one of two conditions (utilitarian loss vs. symbolic loss) and read a scenario involving a service failure in the taxi ride context [64]. In the symbolic loss condition, participants were informed that their needs were ignored, whereas in the utilitarian loss condition, they were told that some taxi service facilities were unavailable when taking a taxi through the YOUJIA platform. The detailed experimental materials are presented in Appendix A.2. Participants were then asked to rate their perceived losses. Following this, they learned that they had made a complaint, which was handled either by a human agent (in the human condition) or by a chatbot (in the chatbot condition). They reviewed a screenshot of the dialogue regarding their interactions with the human agent or chatbot (see Figures A3 and A4 in Appendix A.2). Finally, participants reported their perceived responsiveness, perceived emotional support, and perceived forgiveness on a 7-point scale. Demographic information was also collected. The reliability and validity of the measurement scales were assessed through confirmatory factor analysis (see Table A3 in Appendix C.2).

4.3. Results

Manipulation checks. First, the scenarios were perceived as highly realistic, with scores significantly higher than the scale midpoint. The results confirmed that participants across different groups all rated the experimental scenario as extremely realistic ($M_{symbolic}$ = 5.19, t(259) = 73.92, $p < 0.001$; $M_{utilitarian}$ = 5.07, t(259) = 70.01, $p < 0.001$). Second, the service recovery agent was perceived more as a chatbot in the chatbot group and more as a human in the human group ($M_{chatbot}$ = 1.91, SD = 0.62; M_{human} = 5.72, SD = 0.65; t(518) = −68.5, $p < 0.001$), indicating successful manipulation of the service recovery agent. Regarding consumer loss, participants reported experiencing greater utili-

tarian loss than symbolic loss in the utilitarian loss condition ($M_{symbolic}$ = 4.50, SD = 1.29; $M_{utilitarian}$ = 5.52, SD = 1.29; t(518) = 9.34, $p < 0.01$), while they reported greater symbolic loss than utilitarian loss in the symbolic loss condition ($M_{symbolic}$ = 4.93, SD = 0.84; $M_{utilitarian}$ = 3.26, SD = 1.11; t(518) = −19.4, $p < 0.001$), indicating successful manipulation of consumer loss. Additionally, participants across the four groups showed no significant difference in their current emotions (F(1, 516) = 3.52, $p > 0.05$).

Interaction between service recovery agent and consumer loss. An ANOVA revealed significant main effects of consumer loss ($M_{utilitarian}$ = 3.60, SD = 0.06; $M_{symbolic}$ = 4.21, SD = 0.06; F(1, 516) = 78.33, $p < 0.001$, $\eta^2 = 0.05$) and service recovery agent (M_{human} = 4.04, SD = 0.06; $M_{chatbot}$ = 3.78, SD = 0.06; F(1, 516) = 26.68, $p < 0.001$, $\eta^2 = 0.05$) on consumer forgiveness. The 2 × 2 ANOVA showed a significant interaction between consumer loss and service recovery agent on consumer forgiveness (F(1, 516) = 19.81, $p < 0.001$, $\eta^2 = 0.01$). For utilitarian loss, the service recovery agent had no significant effect on consumer forgiveness (M_{human} = 3.52, SD = 0.97; $M_{chatbot}$ = 3.46, SD = 1.34; F(1, 516) = 0.19, $p > 0.1$). However, for symbolic loss, humans outperformed chatbots in promoting consumer forgiveness (M_{human} = 4.79, SD = 0.92; $M_{chatbot}$ = 3.86, SD = 1.05; F(1, 516) = 42.03, $p < 0.001$). These results replicate the findings from Study 1, supporting H1.

Perceived emotional support. The ANOVA revealed significant main effects of service recovery agent (M_{human} = 4.23, SD = 0.07; $M_{chatbot}$ = 3.46, SD = 0.07; F(1, 516) = 57.40, $p < 0.001$, $\eta^2 = 0.11$) and consumer loss ($M_{utilitarian}$ = 3.48, SD = 0.07; $M_{symbolic}$ = 4.20, SD = 0.07; F(1, 516) = 50.64, $p < 0.001$, $\eta^2 = 0.10$) on perceived emotional support. Additionally, there was a significant interaction between service recovery agent and consumer loss on perceived emotional support (F(1, 516) = 25.93, $p < 0.001$, $\eta^2 = 0.07$; see Figure 3). Specifically, humans were perceived to provide more emotional support than chatbots in the case of symbolic loss (M_{human} = 4.86, SD = 0.97; $M_{chatbot}$ = 3.55, SD = 1.22; F(1, 516) = 76.28, $p < 0.001$). This difference was reduced in the case of utilitarian loss (M_{human} = 3.62, SD = 1.07; $M_{chatbot}$ = 3.35, SD = 1.33; F(1, 516) = 3.31, $p > 0.05$).

Perceived responsiveness. The ANOVA revealed that service recovery agent (M_{human} = 4.16, SD = 0.07; $M_{chatbot}$ = 4.13, SD = 0.07; F(1, 516) = 0.08, $p > 0.05$, $\eta^2 = 0.01$) had no significant effects on perceived responsiveness, while consumer loss ($M_{utilitarian}$ = 4.56, SD = 0.07; $M_{symbolic}$ = 3.73, SD = 0.07; F(1, 516) = 64.05, $p < 0.001$, $\eta^2 = 0.19$) had significant effects on perceived responsiveness. Additionally, there was a significant interaction between service recovery agent and consumer loss on perceived responsiveness (F(1, 516) = 5.39, $p < 0.05$, $\eta^2 = 0.01$; see Figure 4). However, the service recovery agent had no significant effect on perceived responsiveness, regardless of the type of loss (utilitarian loss: M_{human} = 4.70, SD = 1.05; $M_{chatbot}$ = 4.43, SD = 1.09; F(1, 516) = 2.88, $p > 0.05$; symbolic loss: M_{human} = 3.60, SD = 1.52; $M_{chatbot}$ = 3.85, SD = 1.07; F(1, 516) = 2.52, $p > 0.1$).

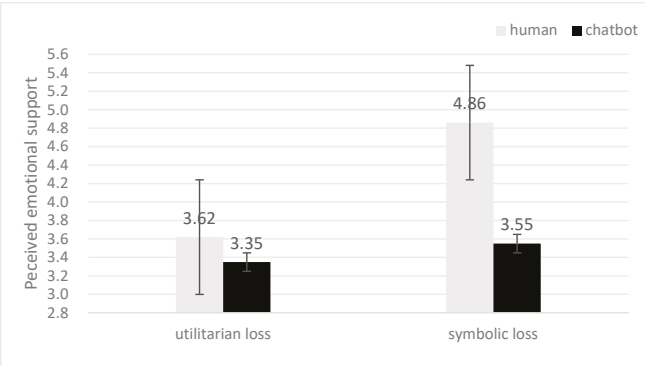


Figure 3. Interaction between service recovery agent and consumer loss on perceived emotional support.

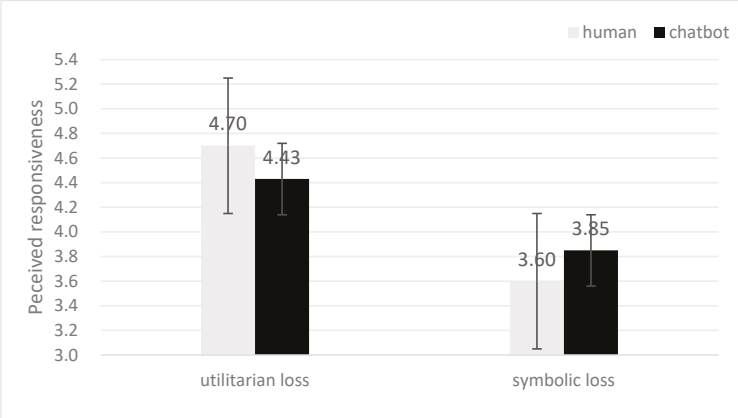


Figure 4. Interaction between service recovery agent and consumer loss on perceived responsiveness.

Moderated mediation analysis. We conducted a moderated mediation analysis using PROCESS Model 8 with service recovery agents as the independent variable [78], consumer forgiveness as the dependent variable, consumer loss as the moderator, and perceived emotional support and perceived responsiveness as the two mediators. The results showed that the indirect effect of service recovery agent on consumer forgiveness through perceived responsiveness was significant in the case of utilitarian loss ($\beta = -0.03$, SE = 0.02, 95% CI = [−0.0784, −0.0022]). However, the indirect effect of service recovery agent on consumer forgiveness through perceived responsiveness was not significant in the case of symbolic loss ($\beta = 0.02$, SE = 0.02, 95% CI = [−0.0048, 0.0745]). Additionally, the indirect effect of service recovery agent on consumer forgiveness through perceived emotional support was not significant in the case of utilitarian loss ($\beta = 0.06$, SE = 0.03, 95% CI = [−0.0020, 0.1307]). In contrast, the indirect effect of service recovery agent on consumer forgiveness through perceived emotional support was significant in the case of symbolic loss ($\beta = 0.28$, SE = 0.07, 95% CI = [0.1520, 0.4224]). These findings indicate that humans are perceived to provide more emotional support than chatbots in the case of symbolic loss, thereby increasing the likelihood of consumer forgiveness. However, in the case of utilitarian loss, both humans and chatbots are perceived as highly responsive, enabling both to effectively enhance consumer forgiveness. These results support H2a and H2b.

5. Study 3

5.1. Pretest

Study 3 aimed to investigate how the interaction between service recovery agent type and consumer loss influences consumer forgiveness, with a particular focus on how this relationship is moderated by service failure severity. To achieve this, a pretest was conducted to validate the manipulation of service failure severity. We first compiled a range of service failure scenarios involving both symbolic and utilitarian losses, sourced from cases reported on a website. Six graduate students were asked to rate the severity of these scenarios, and based on their ratings, we selected the highest-rated scenarios (high-severity) and the lowest-rated scenarios (low-severity) for each type of loss as the experimental stimuli.

A total of 191 participants were recruited for the pretest, with 160 valid responses retained ($M_{age} = 22.8$; 53.42% female). Participants were randomly assigned to different severity conditions within each loss type. The results indicated that, regardless of the consumer loss type, participants in the high-severity condition perceived significantly more severe service failures than those in the low-severity condition ($M_{low} = 4.91$, SD = 1.62; $M_{high} = 5.76$, SD = 1.10; $F(1, 158) = 115.23$, $p < 0.001$). This confirmed that the manipulation

of service failure severity was successful. These experimental stimuli were then used in the formal experiment.

5.2. Design and Participants

Study 3 adopted a 2 (consumer loss: symbolic vs. utilitarian) × 2 (recovery agent: chatbot vs. human) × 2 (service failure severity: low vs. high) between-subjects design to examine whether the interaction between consumer loss and service recovery agent on consumer forgiveness was influenced by service failure severity (H3). A total of 1200 participants were randomly recruited through Wenjuanxing and received a reward of 2 yuan. After excluding invalid responses, 995 valid responses were retained (M_{age} = 23.90; 63.97% female).

5.3. Procedure

Following the procedures used in the previous studies, participants were first asked to report their current emotions and their experience interacting with chatbots. Participants were then randomly assigned to one of four conditions (utilitarian loss vs. symbolic loss; high-severity vs. low-severity) and read a scenario of service failure in the takeaway service context [69]. Participants were informed that they encountered a poor service attitude (in the symbolic loss condition) or missing meals (in the utilitarian loss condition) when ordering from the YOUJIA takeaway platform. In the context of utilitarian loss, the seller forgot to deliver free (paid) items in the low-severity condition (in the high-severity condition). In the symbolic loss context, the merchant impatiently (rudely) agreed to the consumer's request to change the food's taste in the low-severity condition (in the high-severity condition). The detailed experimental materials are shown in Appendix A.3. Participants were then asked to rate their perceived loss and the severity of the service failure. Next, they read that they had made a complaint and that their complaint was handled either by a human agent (in the human condition) or by a chatbot (in the chatbot condition). Finally, participants reported their perceived forgiveness, and demographic information was also collected. The reliability and validity of the measurement scales were assessed through confirmatory factor analysis (see Table A4 in Appendix C.3).

5.4. Results

Manipulation Checks. First, for the realism check, the results confirmed that participants in both groups perceived the experimental scenario as highly realistic ($M_{symbolic}$ = 5.81, t(495) = 16.32, $p < 0.001$; $M_{utilitarian}$ = 5.06, t(498) = 10.86, $p < 0.001$). Second, the service recovery agent was more strongly perceived as a chatbot in the chatbot group and more strongly perceived as a human in the human group ($M_{chatbot}$ = 2.01, SD = 1.16; M_{human} = 5.98, SD = 1.02; t(993) = −26.31, $p < 0.001$). Regarding consumer loss, participants reported greater utilitarian loss than symbolic loss in the utilitarian loss condition ($M_{symbolic}$ = 5.05, SD = 1.27; $M_{utilitarian}$ = 5.27, SD = 0.94; t(993) = −13.10, $p < 0.01$), while they reported greater symbolic loss than utilitarian loss in the symbolic loss condition ($M_{symbolic}$ = 5.46, SD = 1.28; $M_{utilitarian}$ = 4.42, SD = 1.22; t(993) = −13.10, $p < 0.001$), indicating successful manipulation of consumer loss. Service failure severity was perceived as higher in the high-severity condition than in the low-severity condition (M_{high} = 5.25, SD = 1.24; M_{low} = 4.62, SD = 1.30; t(993) = −7.81, $p < 0.001$), confirming the successful manipulation of service failure severity. Additionally, participants across the eight groups showed no significant difference in current emotions (F(1, 987) = 3.06, $p > 0.05$).

Moderating Effect of Service Failure Severity. An ANOVA revealed significant main effects of service failure severity (M_{high} = 3.82, SD = 1.19; M_{low} = 4.21, SD = 1.18; F(1, 987) = 26.66, $p < 0.001$, η^2 = 0.03), consumer loss ($M_{utilitarian}$ = 4.26, SD = 1.13;

$M_{symbolic}$ = 3.77, SD = 1.22; F(1, 987) = 43.97, $p < 0.001$, $\eta^2 = 0.04$), and service recovery agent type (M_{human} = 4.47, SD = 1.07; $M_{chatbot}$ = 3.65, SD = 1.17; F(1, 987) = 131.10, $p < 0.001$, $\eta^2 = 0.12$) on consumer forgiveness. The results also showed a significant interaction between consumer loss, service recovery agent, and service failure severity on consumer forgiveness (F(1, 987) = 4.74, $p < 0.05$, $\eta^2 = 0.01$). In the low-severity condition, the interaction between consumer loss and service recovery agent on consumer forgiveness was significant (F(1, 987) = 14.32, $p < 0.01$, see Figure 5). Specifically, for symbolic loss, humans performed better than chatbots in promoting consumer forgiveness ($M_{chatbot}$ = 3.45, SD = 1.19; M_{human} = 4.57, SD = 0.97; F(1, 987) = 8.61, $p < 0.01$). However, for utilitarian loss, this effect was mitigated ($M_{chatbot}$ = 4.71, SD = 1.00; M_{human} = 4.47, SD = 1.01; F(1, 987) = 5.73, $p > 0.05$). In the high-severity condition, the interaction between consumer loss and service recovery agent on consumer forgiveness was not significant (F(1, 987) = 0.45, $p > 0.05$, see Figure 6). Regardless of the loss type, humans performed better than chatbots in promoting consumer forgiveness in the high-severity condition (symbolic loss: $M_{chatbot}$ = 3.29, SD = 1.06; M_{human} = 4.12, SD = 1.14; F(1, 987) = 42.93, $p < 0.001$; utilitarian loss: $M_{chatbot}$ = 3.69, SD = 1.10; M_{human} = 4.48, SD = 1.09; F(1, 987) = 32.64, $p < 0.001$).

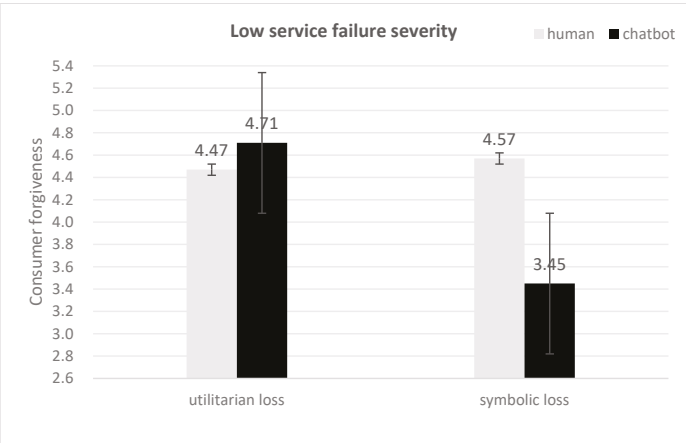


Figure 5. Interaction between consumer loss and recovery agent in the low-severity condition.

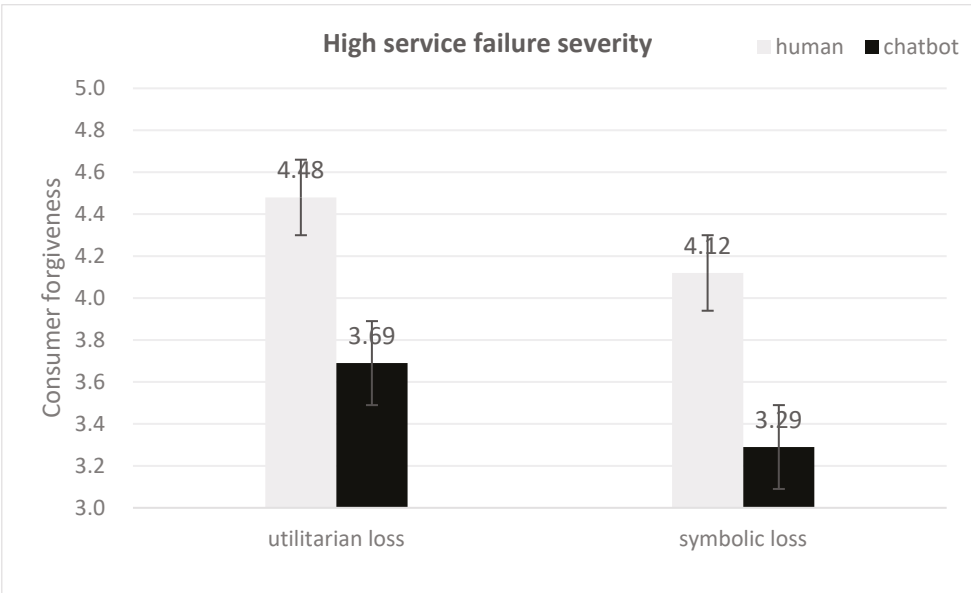


Figure 6. Interaction between consumer loss and recovery agent in the high-severity condition.

6. Discussion

6.1. General Discussion

With the advancement of machine learning technologies, an increasing number of companies are deploying chatbots as service recovery agents. As a result, consumers are progressively turning to chatbots to resolve complaints arising from service failures. This practice has raised an important question: Can chatbots address service failures as effectively as human agents? To investigate this, three studies were conducted to examine how the interplay between service recovery agents, consumer loss, and service failure severity influences consumer forgiveness in the context of service failures.

The impact of service recovery agents on consumer forgiveness was found to be contingent upon the type of consumer loss. In the case of symbolic loss, addressing consumer complaints through a human agent is more effective in fostering forgiveness than handling them through a chatbot. However, this effect is diminished in the context of utilitarian loss. To explore the underlying mechanisms of this phenomenon, we examined consumer behavior in these contexts. Specifically, when consumers experience symbolic loss, they seek emotional support [57]. Existing literature suggests that chatbot interactions tend to reduce consumers' emotional experiences [79]. In contrast, humans are better equipped to provide emotional support, making them more effective in promoting consumer forgiveness in the case of symbolic loss. On the other hand, when consumers experience utilitarian loss, they prioritize efficient problem-solving. From a utilitarian perspective, consumers expect recovery services to be accurate, responsive, and compatible [80]. Both human agents and chatbots exhibit high responsiveness, allowing them to effectively enhance consumer forgiveness in cases of utilitarian loss.

We also found that the interaction between service recovery agents and consumer loss is influenced by the severity of the service failure. In the case of high-severity service failures, consumers respond more favorably to recovery services provided by humans compared to those provided by chatbots, regardless of the type of loss. However, for low-severity service failures, humans are more effective than chatbots in addressing service failures and promoting consumer forgiveness, particularly in the case of symbolic loss.

6.2. Theoretical Contributions

First, our study contributes to the existing literature on service failures and service recovery by examining how service recovery agents influence consumer forgiveness across different types of consumer loss. While much of the existing research on service failures primarily focuses on exploring recovery strategies, such as apologies, compensation, and explanations [6,7], our paper enriches this body of work by proposing that the use of chatbots to handle consumer complaints can serve as an effective method for promoting service recovery.

Second, the current study makes a significant contribution by extending research on human-AI interaction to service failure scenarios. Recent studies have focused on consumers' acceptance of AI-driven services, marking a prominent stream of research in the field of human-AI interaction [81,82]. Numerous studies have highlighted consumers' varied responses to human-provided versus AI-provided services. Building on these findings, our study explores consumer reactions to complaints handled by chatbots and human agents in service failure situations. While some research has examined the application of AI in managing service failures [83,84], few studies address the interactive effect of service recovery agents and consumer loss on consumer forgiveness. Our results show that chatbots only foster consumer forgiveness in utilitarian loss contexts, whereas they may have a counterproductive effect in symbolic loss contexts. In contrast, human

agents promote consumer forgiveness in both symbolic and utilitarian loss contexts. Thus, our study identifies an important boundary condition under which consumers respond differently to human-provided and chatbot-provided services.

Third, this study identifies the mechanisms—perceived emotional support and perceived responsiveness—through which the interaction between recovery agent type and consumer loss influences consumer forgiveness in service failure contexts. While prior studies have identified several mediators like perceived justice and cognitive appraisal, these mediators rarely explain why agent type matters for recovery effectiveness across loss contexts. Perceived justice, for example, treats recovery agents as interchangeable vehicles for delivering fair outcomes, while cognitive appraisal focuses on loss severity rather than agent characteristics. In contrast, perceived emotional support and perceived responsiveness highlight that chatbots and humans differ in their ability to promote consumer forgiveness. This integration clarifies why loss type shapes different agents' effectiveness in addressing service failures. We found that both chatbots and humans are capable of providing timely responses, which fosters consumer forgiveness in utilitarian loss contexts. However, humans are more adept at offering emotional support, thereby promoting consumer forgiveness in symbolic loss contexts. These findings contribute to a deeper understanding of why consumers respond differently to recovery services provided by chatbots and human agents under varying loss conditions.

Finally, by incorporating cognitive appraisal processes (controllability, stability, intentionality), our study identifies service failure severity as a contextual moderator that influences the interaction between recovery agent type and consumer loss. In low-severity service failures, recovery services provided by a human agent (vs. a chatbot) are more effective in facilitating consumer forgiveness, particularly in symbolic loss contexts. However, in high-severity failures, which trigger more intense cognitive appraisals, consumers exhibit more positive responses to recovery services provided by a human agent, regardless of the loss type.

6.3. Managerial Insights

Our findings offer actionable and nuanced managerial implications for service-oriented firms seeking to optimize service recovery strategies, balance cost efficiency, and enhance consumer forgiveness.

Firms should establish a dynamic service recovery routing system that prioritizes the alignment between recovery agents (human vs. chatbot), consumer loss type, and service failure severity. For symbolic loss scenarios, human agents remain irreplaceable. As our results confirm, humans outperform chatbots in providing perceived emotional support, which is a key mediator of forgiveness for symbolic loss. Managers should ensure that complaints involving terms such as "disrespect," "rude," or "humiliating" are automatically escalated to human agents. For utilitarian loss scenarios, chatbots serve as a cost-effective and equally effective alternative to humans. Our study shows that both agents achieve comparable levels of perceived responsiveness, the primary mediator of forgiveness for utilitarian losses. Firms can leverage chatbots to handle consumers' complaints, as they offer 24/7 availability and lower operational costs while meeting consumers' core demand for timely, practical solutions [15,64]. Therefore, technicians can develop classification models that automatically distinguish between utilitarian and symbolic losses by analyzing complaint text, emotional expressions, and loss descriptions [85]. Moreover, to mitigate misclassification risks, chatbots must include a prominent "request human assistance" option, which allows consumers to override automated routing if they perceive unmet emotional needs. Regarding service failure severity, human agents are mandatory for high-severity incidents, regardless of loss type. High-severity failures trigger intense

cognitive appraisals, with consumers perceiving the incident as less controllable, more stable, or intentional, which amplifies the need for sincere, empathetic recovery [26,28]. Human agents excel at addressing these complex emotional and cognitive needs, whereas chatbots' perceived lack of emotional intelligence may exacerbate dissatisfaction [11,75]. For low-severity failures, chatbots remain viable for utilitarian losses, while humans are still preferred for symbolic losses to maintain relational trust.

While chatbots cannot fully replicate human emotional intelligence, firms can optimize their design to improve performance in symbolic loss contexts, thus reducing the effectiveness gap with humans. First, anthropomorphism and emotional scripting are critical. Research shows that anthropomorphic design enhances perceived humanness and empathy in chatbots [23,45]. For symbolic loss scenarios, chatbot scripts can be tailored to include empathetic expressions, emoticons, and acknowledgment of emotional distress [25]. Such scripts align with interactional justice principles, addressing the core injustice of symbolic losses. Second, integrating humor and emotional reinforcement can further enhance perceived emotional support. Studies have shown that humorous or self-mocking communication in service recovery reduces negative emotions and increases forgiveness [8,25]. Chatbots can be programmed to use lighthearted humor in low-to-moderate severity symbolic loss scenarios. Additionally, chatbots can leverage facial emojis to convey empathy, as facial emojis have been shown to increase service recovery satisfaction compared to non-facial emojis [48]. Third, hybrid human-chatbot models offer a middle ground for symbolic loss contexts. Chatbots can serve as first-line responders to triage complaints, gather basic information, and provide initial apologies. If the chatbot detects emotional cues or receives a high-severity rating from the consumer, the interaction is seamlessly transferred to a human agent within 1–2 exchanges [44]. This model combines chatbots' efficiency in routine tasks with humans' strength in emotional recovery and minimizes consumer frustration while controlling costs.

Firms should adopt a data-driven approach to refine service recovery systems, using consumer feedback and interaction data to improve agent routing, chatbot performance, and training programs. First, feedback loops can be implemented post-recovery to measure perceived emotional support, responsiveness, and forgiveness. For example, consumers can be asked short surveys to identify gaps, such as chatbots failing to recognize symbolic losses or human agents lacking empathy. This data can be used to retrain chatbot algorithms and enhance human agent training. Second, predictive analytics can help anticipate failure severity and consumer responses. By analyzing factors such as complaint content, consumer history, and contextual details, firms can proactively route high-risk complaints to human agents to prevent dissatisfaction from escalating [2,86].

6.4. Limitations and Future Research

The current study has several limitations in three aspects, including experimental design, the robustness of the results, and the generalizability of the findings. First, in terms of experimental design, scenario-based experiments were employed to test our hypotheses, which may limit the external validity of our findings. Future research could conduct field experiments to validate and generalize our results in real-world settings. In addition, while the pretest and Study 3 confirmed statistically significant differences between the intended "low-severity" and "high-severity" conditions, both conditions clustered within the mid-to-high range of the severity scale. This suggests that the "low-severity" scenarios in the current study may not represent minor, easily overlooked service failures. Future research could develop more finely differentiated severity manipulations that can span the entire severity spectrum. Specifically, low-severity service failures could be defined as minimal non-material inconvenience—for example, an online retail platform sending

a non-essential order confirmation email with a minor typo. In contrast, high-severity failures should involve substantial tangible losses or threats to essential consumer interests, such as the leakage of consumer personal information by hotels.

Second, in terms of the robustness of the results, individual differences, such as technology readiness, social anxiety, and anthropomorphism tendencies, may influence preferences for service recovery agents. Future research should investigate the impact of these individual differences on consumers' preferences for human agents versus chatbots. Additionally, our study draws upon previous findings to suggest that anthropomorphic chatbots may perform as effectively as human agents in promoting consumer forgiveness for service failures. However, we have not empirically tested this assertion. Future research could manipulate the level of anthropomorphism in chatbots by using human-like avatars and names and examine the effects of chatbot anthropomorphism on consumer responses in service failure contexts.

Third, in terms of the generalizability of the findings, our study focused solely on data collected from China, which may limit the generalizability of our findings. In collectivist cultures like China, where social harmony and relational respect are highly valued [50], consumers may have heightened expectations for emotional support in symbolic loss contexts. In contrast, individualistic cultures may prioritize individual autonomy and efficiency, potentially leading to greater acceptance of chatbots in symbolic loss situations if they provide prompt and transparent recovery [17]. Therefore, future research should collect data from different cultural backgrounds (such as Western countries vs. China) and investigate the influence of cultural backgrounds on consumer responses in service failure contexts. Moreover, the perception of loss type and severity may vary across different experimental situations. For example, in the medical service scenario, consumers may have a particularly strong perception of the severity of the failure of doctors' services. Therefore, our research findings may not be applicable to other service domains. Future research could consider verifying consumers' preferences for service recovery agents in high-risk service domains (e.g., medical diagnosis and financial investment) when they encounter service failures.

Author Contributions: Conceptualization, L.F. and C.W.; methodology, S.L.; software, C.W.; validation, L.F., S.L. and C.W.; formal analysis, S.L.; investigation, C.W.; resources, S.L.; data curation, X.Z.; writing—original draft preparation, S.L.; writing—review and editing, X.Z.; visualization, S.L.; supervision, L.F.; project administration, X.Z.; funding acquisition, L.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [the Natural Science Foundation of Shandong Province] grant number [ZR2022MG015], [National Natural Science Foundation of China] grant number [72271082, 72571086], and [Natural Science Foundation of Anhui Province] grant number [2408085J041].

Institutional Review Board Statement: This study was exempt from ethical review and approval, while strictly abiding by the ethical guidelines outlined in the Declaration of Helsinki. The Institutional Review Board of Wuhan University of Technology waived the requirement for ethical review and approval in line with existing Chinese regulations. The research utilized fully anonymized data without personal identifiers such as names or national ID numbers. This meets the provisions of Article 16 of the Ethical Review Measures for Biomedical Research Involving Humans (National Health Commission Order No. 11, 2016), which grants exemption from ethical approval for studies using non-traceable anonymized data. Additionally, Article 4 of China's Personal Information Protection Law (2021) legally classifies anonymized data outside the scope of personal information, freeing it from restrictions under personal data protection rules.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available on the request from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Experimental Stimuli Used in Study 1

Consider the process of checking into a "YOUJIA" hotel. Upon arrival at the reception desk, you request the assistance of the attendant to convey your luggage to your room.

The service staff had a cold and indifferent attitude and reluctantly helped you deliver your luggage to your room. You were dissatisfied with the service experience and complained on the YOUJIA online platform [symbolic loss condition].

You make a complaint and it is handled by a human/chatbot [human condition/chatbot condition] as shown in the picture below. The human/chatbot provides an acceptable solution which resolves your query.

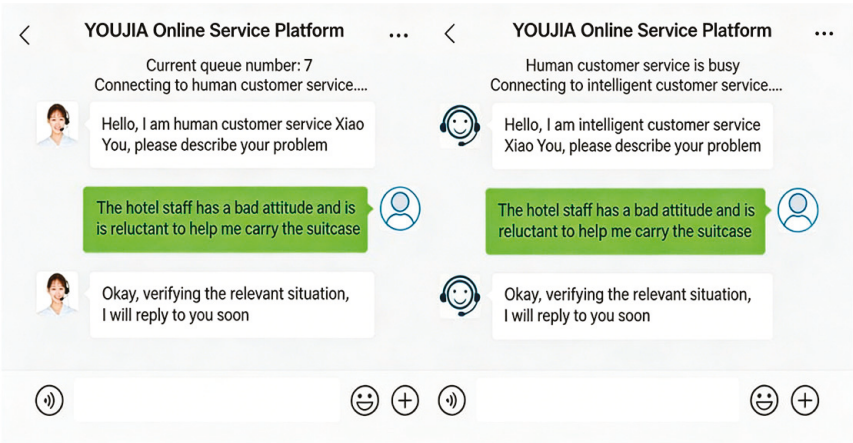


Figure A1. Symbolic loss material in Study 1. ((**Left**) human condition; (**Right**) chatbot condition).

The service staff helped you deliver your luggage to your room. However, you found that the exterior of your suitcase was damaged. You were dissatisfied with the service experience and complained on the YOUJIA online platform [utilitarian loss condition].

You make a complaint and it is handled by a human/chatbot [human condition/chatbot condition] as shown in the picture below. The human/chatbot provides an acceptable solution which resolves your query.

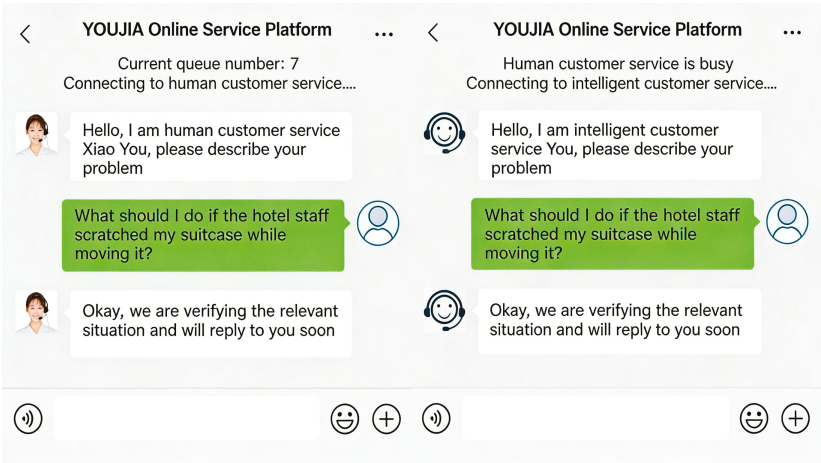


Figure A2. Utilitarian loss material in Study 1. ((**Left**) human condition; (**Right**) chatbot condition).

Appendix A.2. Experimental Stimuli Used in Study 2

Consider the process of taking a taxi with complimentary Wi-Fi and air conditioning at the "YOUJIA" transportation platform.

Upon entering the vehicle, you found that the air conditioning was not turned on and then requested the driver to turn it on. However, rather than complying with this request, the driver just opened the car window instead of turning on the air conditioning. And then you made a second request. The driver agreed to your request and turned on the air conditioning. Upon arriving at the airport, you complained to the service provider via an app chat interface [symbolic loss condition].

You make a complaint and it is handled by a human/chatbot [human condition/chatbot condition] as shown in the picture below. The human/chatbot provides an acceptable solution which resolves your query.

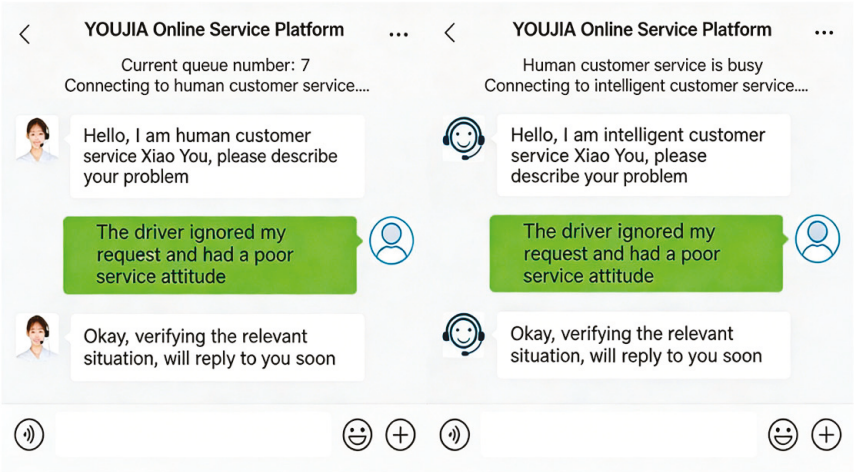


Figure A3. Symbolic loss material in Study 2. ((**Left**) the human condition; (**Right**) chatbot condition).

Upon entering the vehicle, you found that there were no air conditioning or WiFi facilities in the taxi. Upon arriving at the airport, you complained to the service provider via an app chat interface. [utilitarian loss condition]

You make a complaint and it is handled by a human/chatbot [human condition/chatbot condition] as shown in the picture below. The human/chatbot provides an acceptable solution which resolves your query.

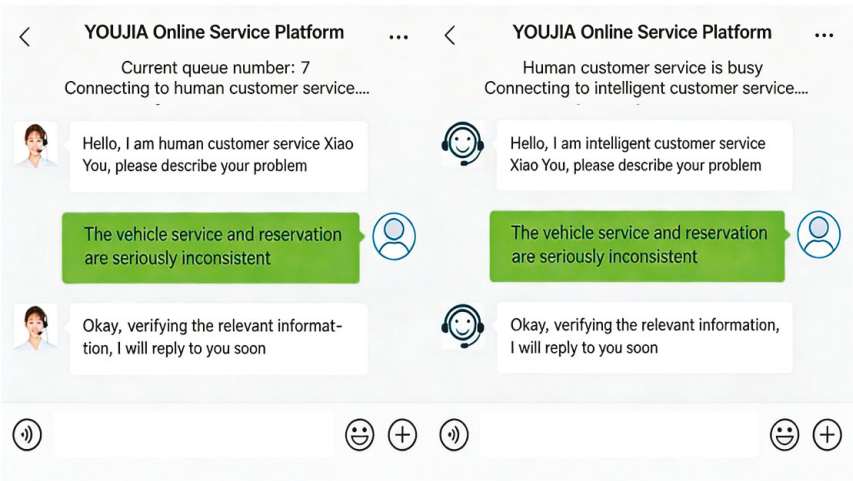


Figure A4. Utilitarian loss material in Study 2. ((**Left**) the human condition; (**Right**) chatbot condition).

Appendix A.3. Experimental Stimuli Used in Study 3

Consider the process of ordering a takeaway at the "YOUJIA" takeaway platform [utilitarian loss condition].

Upon receipt of the takeout order, you found that the merchant did not deliver the free drink they promised. Given the unsatisfactory service, you complained to the YOUJIA takeaway platform [low-level service failure severity condition].

You make a complaint and it is handled by a human/chatbot [human condition/chatbot condition] as shown in the picture below. The human/chatbot provides an acceptable solution which resolves your query.

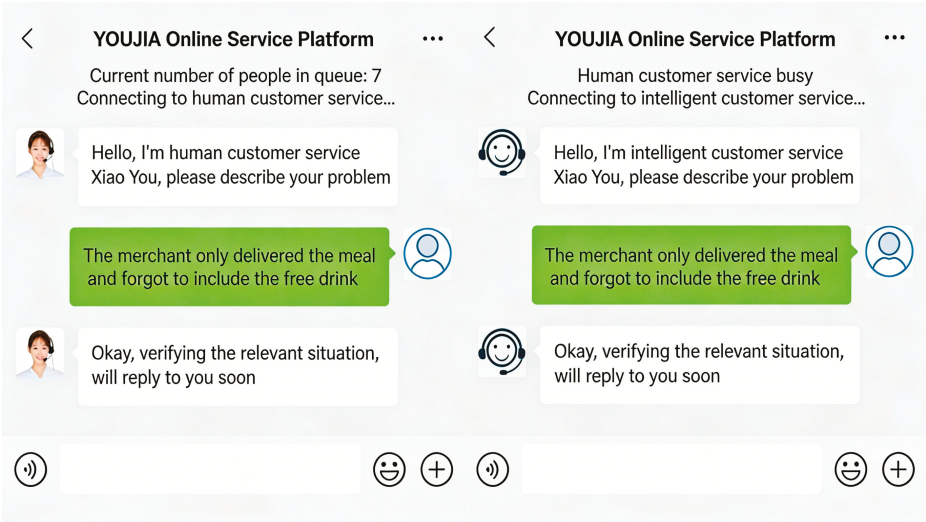


Figure A5. Utilitarian loss × low service failure severity material in Study 3. ((**Left**) the human condition; (**Right**) chatbot condition).

Upon receipt of the takeout order, you found that the merchant did not deliver the free drink they promised and the meal you paid. Given the unsatisfactory service, you complained to the YOUJIA takeaway platform [high-level service failure severity condition].

You make a complaint and it is handled by a human/chatbot [human condition/chatbot condition] as shown in the picture below. The human/chatbot provides an acceptable solution which resolves your query.

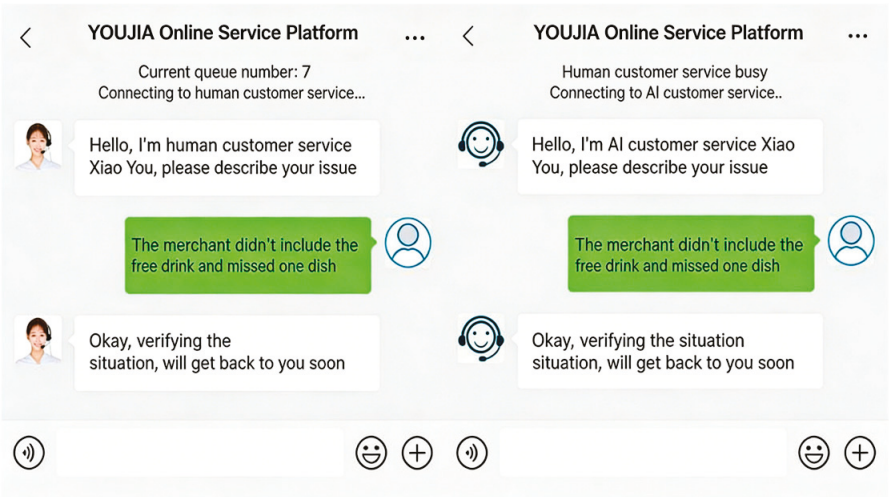


Figure A6. Utilitarian loss × high service failure severity material in Study 3. ((**Left**) the human condition; (**Right**) chatbot condition).

Consider the process of making a phone call to the restaurant to ask to make the food you ordered lighter at the "YOUJIA" takeaway platform. [symbolic loss condition]

You called the merchant to request to make your food lighter. The merchant impatiently agreed to my request and just said "ok" and then hung up immediately. Given the unsatisfactory service, you complained to the YOUJIA takeaway platform [low-level service failure severity condition].

You make a complaint and it is handled by a human/chatbot [human condition/chatbot condition] as shown in the picture below. The human/chatbot provides an acceptable solution which resolves your query.

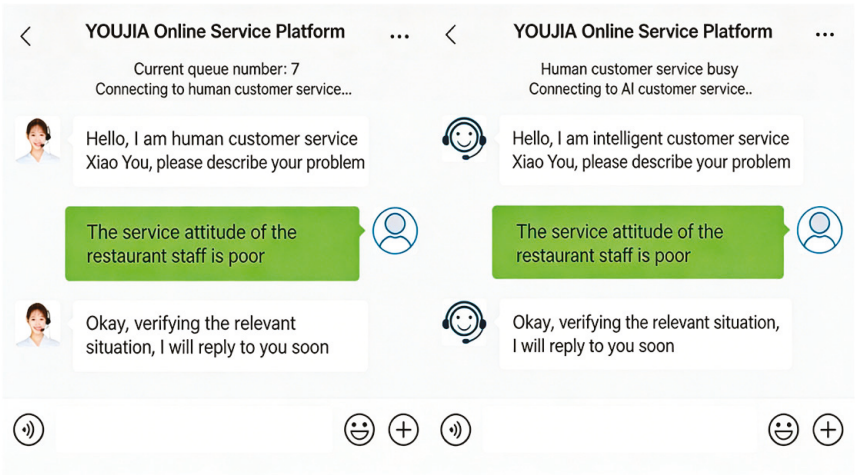


Figure A7. Symbolic loss × low service failure severity material in Study 3. ((**Left**) the human condition; (**Right**) chatbot condition).

You called the merchant to request to make your food lighter. The merchant impatiently agreed to my request. And the manager said "Can you fill in the notes yourself next time? I am too busy!" and hung up immediately. Given the unsatisfactory service, you complained to the YOUJIA takeaway platform [high-level service failure severity condition].

You make a complaint and it is handled by a human/chatbot [human condition/chatbot condition] as shown in the picture below. The human/chatbot provides an acceptable solution which resolves your query.

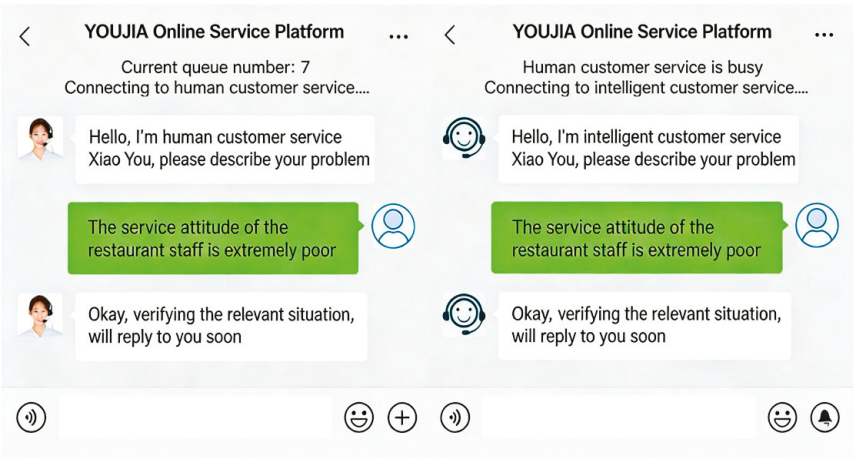


Figure A8. Symbolic loss × high service failure severity material in Study 3. ((**Left**) the human condition; (**Right**) chatbot condition).

Appendix B. Items

Table A1. Items.

Construct	Source	Items
Utilitarian loss	Smith et al., 1999 [77]	(1) The hotel service didn't meet my original expectation (2) it delayed my time (3) it didn't strike me as good value for money (4) it resulted in wasted money
Symbolic loss	Smith et al., 1999 [77]	(1) I perceived a lack of respect from the hotel staff (2) The hotel staff was very unfriendly (3) The hotel staff's actions undermined my dignity (4) I felt unfairly treated during this service
Consumer forgiveness	Fedorikhin et al., 2008 [87]	(1) I will understand the hotel service mistakes (2) I will forgive the errors of the hotel service (3) I choose to forgive the hotel (4) I will continue to trust the hotel
Emotional state	Townsen & Sood, 2012 [76]	I feel: (1) Sad-Happy (2) In a bad mood-a good mood (3) Irritable-Pleased (4) Depressed-Cheerful
Perceived responsiveness	Chen et al., 2024 [61]	(1) The agent can quickly respond to my inquiries (2) When I require assistance from the agent, I consistently receive timely responses (3) The agent is always available online (4) I communicate with the agent very efficiently
Perceived emotional support	Gelbrich et al., 2021 [51]	Feedback from the agent makes me feel . . . (1) emotional support (2) emotional reinforcement (3) comfort (4) Assistance
Service failure severity	Ho et al., 2020 [11]	The above-mentioned service failure that happened to me was . . . (1) Minor problems-Major problems (2) Small inconveniences Big inconveniences (3) Minor aggravation Major aggravation

Appendix C.

Appendix C.1. Confirmation Factor Analysis of Study 1

Table A2. Confirmation Factor Analysis of Study 1.

Variable	Item	Factor Loading	Cronbach's α	CR	AVE
Symbolic loss	SL1	0.900	0.904	0.758	0.926
	SL2	0.930			
	SL3	0.804			
	SL4	0.844			

Table A2. *Cont.*

Variable	Item	Factor Loading	Cronbach's α	CR	AVE
Utilitarian loss	UL1	0.808	0.848	0.692	0.899
	UL2	0.772			
	UL3	0.856			
	UL4	0.886			
Consumer forgiveness	CF1	0.886	0.900	0.769	0.930
	CF2	0.901			
	CF3	0.856			
	CF4	0.865			

Appendix C.2. Confirmation Factor Analysis of Study 2

Table A3. Confirmation Factor Analysis of Study 2.

Variable	Item	Factor Loading	Cronbach's α	CR	AVE
Symbolic loss	SL1	0.873	0.861	0.906	0.707
	SL2	0.818			
	SL3	0.824			
	SL4	0.847			
Utilitarian loss	UL1	0.899	0.902	0.932	0.774
	UL2	0.816			
	UL3	0.889			
	UL4	0.913			
Consumer forgiveness	CF1	0.877	0.917	0.941	0.801
	CF2	0.907			
	CF3	0.915			
	CF4	0.880			
Perceived responsiveness	PR1	0.843	0.856	0.878	0.645
	PR2	0.869			
	PR3	0.787			
	PR4	0.703			
Perceived emotional support	PE1	0.904	0.936	0.955	0.842
	PE2	0.920			
	PE3	0.917			
	PE4	0.929			

Appendix C.3. Confirmation Factor Analysis of Study 3

Table A4. Confirmation Factor Analysis of Study 3.

Variable	Item	Factor Loading	Cronbach's α	CR	AVE
Symbolic loss	SL1	0.910	0.909	0.936	0.786
	SL2	0.926			
	SL3	0.858			
	SL4	0.851			
Utilitarian loss	UL1	0.775	0.842	0.895	0.681
	UL2	0.799			
	UL3	0.882			
	UL4	0.841			
Consumer forgiveness	CF1	0.865	0.890	0.924	0.754
	CF2	0.898			
	CF3	0.897			
	CF4	0.810			

Table A4. *Cont.*

Variable	Item	Factor Loading	Cronbach's α	CR	AVE
Service failure severity	SF1	0.838	0.841	0.904	0.759
	SF2	0.881			
	SF3	0.893			

References

1. Migacz, S.J.; Zou, S.; Petrick, J.F. The "Terminal" Effects of Service Failure on Airlines: Examining Service Recovery with Justice Theory. *J. Travel Res.* **2018**, *57*, 83–98. [CrossRef]
2. Owusu, P.; Li, Z.; Mensah, I.A.; Omari-Sasu, A.Y. Consumer response to E-commerce service failure: Leveraging repurchase intentions through strategic recovery policies. *J. Retail. Consum. Serv.* **2025**, *82*, 104137. [CrossRef]
3. Liu, Y.; Cheng, P.; Ouyang, Z. How trust mediate the effects of perceived justice on loyalty: A study in the context of automotive recall in China. *J. Retail. Consum. Serv.* **2021**, *58*, 102322. [CrossRef]
4. Wang, X.; Zhang, Q. Does online service failure matter to offline customer loyalty in the integrated multi-channel context? The moderating effect of brand strength. *J. Serv. Theory Pract.* **2018**, *28*, 774–806. [CrossRef]
5. Grégoire, Y.; Tripp, T.M.; Legoux, R. When customer love turns into lasting hate: The effects of relationship strength and time on customer revenge and avoidance. *J. Mark.* **2009**, *73*, 18–32. [CrossRef]
6. Gelbrich, K.; Roschk, H. A meta-analysis of organizational complaint handling and customer responses. *J. Serv. Res.* **2011**, *14*, 24–43. [CrossRef]
7. Park, J.; Yoo, J.W.; Cho, Y.-E.; Park, H. Examining the impact of service robot communication styles on customer intimacy following service failure. *J. Retail. Consum. Serv.* **2023**, *75*, 103511. [CrossRef]
8. Zhang, Y.; Huang, J.; Pang, Q. Turning setbacks into smiles: Exploring the role of self-mocking strategies in consumers' recovery satisfaction after e-commerce service failures. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 183. [CrossRef]
9. Harrison-Walker, L.J. The critical role of customer forgiveness in successful service recovery. *J. Bus. Res.* **2019**, *95*, 376–391. [CrossRef]
10. Muhammad, L.; Gul, E.R. Mediating role of customer forgiveness between perceived justice and satisfaction. *J. Retail. Consum. Serv.* **2020**, *52*, 101886. [CrossRef]
11. Ho, T.H.; Tojib, D.; Tsarenko, Y. Human staff vs. service robot vs. fellow customer: Does it matter who helps your customer following a service failure incident? *Int. J. Hosp. Manag.* **2020**, *87*, 102501. [CrossRef]
12. Majeed, S.; Kim, W.G.; Nimri, R. Conceptualizing the role of virtual service agents in service failure recovery: Guiding insights. *Int. J. Hosp. Manag.* **2024**, *123*, 103889. [CrossRef]
13. Peng, Y.; Wang, Y.; Li, J.; Yang, Q. Impact of AI-oriented live-streaming e-commerce service failures on consumer disengagement—Empirical evidence from China. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 1580–1598. [CrossRef]
14. Kale, A.; Nguyen, T.; Harris, F.C.H., Jr.; Li, C.H.; Zhang, J.Y.; Ma, X.G. Provenance documentation to enable explainable and trustworthy AI: A literature review. *Data Intell.* **2023**, *5*, 139–162. [CrossRef]
15. Pizzi, G.; Scarpi, D.; Pantano, E. Artificial intelligence and the new forms of interaction: Who has the control when interacting with a chatbot? *J. Bus. Res.* **2021**, *129*, 878–890. [CrossRef]
16. Li, Y.; Li, Y.; Chen, Q.; Chang, Y. Humans as teammates: The signal of human–AI teaming enhances consumer acceptance of chatbots. *Int. J. Inf. Manag.* **2024**, *76*, 102771. [CrossRef]
17. Longoni, C.; Cian, L. Artificial intelligence in utilitarian vs. hedonic contexts: The "word-of-machine" effect. *J. Mark.* **2022**, *86*, 91–108. [CrossRef]
18. Markovitch, D.G.; Stough, R.A.; Huang, D. Consumer reactions to chatbot versus human service: An investigation in the role of outcome valence and perceived empathy. *J. Retail. Consum. Serv.* **2024**, *79*, 103847. [CrossRef]
19. Castelo, N.; Bos, M.W.; Lehmann, D.R. Task-dependent algorithm aversion. *J. Mark. Res.* **2019**, *56*, 809–825. [CrossRef]
20. Du, J.; Fan, X. The impact of customer's loss and emotion on recovery expectation and complaining intention under the service failure circumstance. *Nankai Bus. Rev.* **2007**, *10*, 4–10.
21. Zhang, W.; Yang, B.; Yang, M.; Xie, L.; Zheng, Y. Sincerity or incompetence? The double-edged sword effect of sadness expression in chatbots service failure. *J. Retail. Consum. Serv.* **2025**, *86*, 104325. [CrossRef]
22. Liu, D.; Lv, Y.; Huang, W. How do consumers react to chatbots' humorous emojis in service failures. *Technol. Soc.* **2023**, *73*, 102244. [CrossRef]
23. Zhang, J.; Lu, X.; Zheng, W.; Wang, X. It's better than nothing: The influence of service failures on user reusage intention in AI chatbot. *Electron. Commer. Res. Appl.* **2024**, *67*, 101421. [CrossRef]
24. Crolic, C.; Thomaz, F.; Hadi, R.; Stephen, A.T. Blame the bot: Anthropomorphism and anger in customer–chatbot interactions. *J. Mark.* **2022**, *86*, 132–148. [CrossRef]

25. Xie, C.; Zhu, J.; Xie, Y.; Liang, C. Apologizing with a smile or crying face? Exploring the impact of emoji types on customer forgiveness within chatbots service recovery. *Electron. Commer. Res. Appl.* **2025**, *70*, 101488. [CrossRef]
26. Joireman, J.; Grégoire, Y.; Devezer, B.; Tripp, T.M. When do customers offer firms a 'second chance' following a double deviation? the impact of inferred firm motives on customer revenge and reconciliation. *J. Retail.* **2013**, *89*, 315–337. [CrossRef]
27. Li, Y.; Zhang, C.; Fang, S. Can beauty save service failures? The role of recovery employees' physical attractiveness in the tourism industry. *J. Bus. Res.* **2022**, *141*, 100–110. [CrossRef]
28. Kuang, D.; Ma, B.; Wang, H. The relative impact of advertising and referral reward programs on the post-consumption evaluations in the context of service failure. *J. Retail. Consum. Serv.* **2022**, *65*, 102849. [CrossRef]
29. Tih, S.; Wong, K.-K.; Lynn, G.S.; Reilly, R.R. Prototyping, customer involvement, and speed of information dissemination in new product success. *J. Bus. Ind. Mark.* **2016**, *31*, 437–448. [CrossRef]
30. Akarsu, T.N.; Marvi, R.; Foroudi, P. Service failure research in the hospitality and tourism industry: A synopsis of past, present and future dynamics from 2001 to 2020. *Int. J. Contemp. Hosp. Manag.* **2023**, *35*, 186–217. [CrossRef]
31. Sann, R.; Lai, P.-C. Understanding homophily of service failure within the hotel guest cycle: Applying NLP-aspect-based sentiment analysis to the hospitality industry. *Int. J. Hosp. Manag.* **2020**, *91*, 102678. [CrossRef]
32. Yang, W.; Mattila, A.S. The role of tie strength on consumer dissatisfaction responses. *Int. J. Hosp. Manag.* **2012**, *31*, 399–404. [CrossRef]
33. Kim, H.; So, K.K.F. The evolution of service failure and recovery research in hospitality and tourism: An integrative review and future research directions. *Int. J. Hosp. Manag.* **2023**, *111*, 103457. [CrossRef]
34. Grönroos, C. A service quality model and its marketing implications. *Eur. J. Mark.* **1984**, *18*, 36–44. [CrossRef]
35. Sparks, B.A.; Browning, V. Complaining in cyberspace: The motives and forms of hotel guests' complaints online. *J. Hosp. Mark. Manag.* **2010**, *19*, 797–818. [CrossRef]
36. Joireman, J.; Grégoire, Y.; Tripp, T.M. Customer forgiveness following service failures. *Curr. Opin. Psychol.* **2016**, *10*, 76–82. [CrossRef]
37. Habib, M.D.; Attri, R.; Salam, M.A.; Yaqub, M.Z. Retail consumers' conundrum: An in-depth qualitative study navigating the motivations and aversion of chatbots. *J. Retail. Consum. Serv.* **2025**, *82*, 104147. [CrossRef]
38. Dietvorst, B.J.; Simmons, J.P.; Massey, C. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *J. Exp. Psychol. Gen.* **2015**, *144*, 114–126. [CrossRef]
39. Garvey, A.M.; Kim, T.; Duhachek, A. Bad news? send an AI. good news? send a human. *J. Mark.* **2021**, *87*, 10–25. [CrossRef]
40. Gunaratne, J.; Zalmanson, L.; Nov, O. The persuasive power of algorithmic and crowdsourced advice. *J. Manag. Inf. Syst.* **2018**, *35*, 1092–1120. [CrossRef]
41. Yalcin, G.; Lim, S.; Puntoni, S.; van Osselaer, S.M.J. Thumbs Up or Down: Consumer Reactions to Decisions by Algorithms Versus Humans. *J. Mark. Res.* **2021**, *59*, 696–717. [CrossRef]
42. Logg, J.M.; Minson, J.A.; Moore, D.A. Algorithm appreciation: People prefer algorithmic to human judgment. *Organ. Behav. Hum. Decis. Process.* **2019**, *151*, 90–103. [CrossRef]
43. Mahmud, H.; Islam, A.K.M.N.; Luo, X.; Mikalef, P. Decoding algorithm appreciation: Unveiling the impact of familiarity with algorithms, tasks, and algorithm performance. *Decis. Support Syst.* **2024**, *179*, 114168. [CrossRef]
44. Sands, S.; Campbell, C.; Plangger, K.; Pitt, L. Buffer bots: The role of virtual service agents in mitigating negative effects when service fails. *Psychol. Mark.* **2022**, *39*, 2039–2054. [CrossRef]
45. Shams, G.; Kim, K.K.; Kim, K. Enhancing service recovery satisfaction with chatbots: The role of humor and informal language. *Int. J. Hosp. Manag.* **2024**, *120*, 14. [CrossRef]
46. Agnihotri, A.; Bhattacharya, S. Chatbots' effectiveness in service recovery. *Int. J. Inf. Manag.* **2024**, *76*, 102679. [CrossRef]
47. Xie, Y.; Zhou, P.; Liang, C.; Zhao, S.; Lu, W. Affiliative or self-defeating? exploring the effect of humor types on customer forgiveness in the context of ai agents' service failure. *J. Bus. Res.* **2025**, *194*, 115381. [CrossRef]
48. Cao, Z.; Yu, K. To facial or not to facial? From emoji to empathy in shaping customer satisfaction with chatbot service recovery. *J. Retail. Consum. Serv.* **2026**, *89*, 104633. [CrossRef]
49. Xing, X.; Song, M.; Duan, Y.; Mou, J. Effects of different service failure types and recovery strategies on the consumer response mechanism of chatbots. *Technol. Soc.* **2022**, *70*, 102049. [CrossRef]
50. Hook, J.N.; Worthington, E.L.; Utsey, S.O. Collectivism, Forgiveness, and Social Harmony. *Couns. Psychol.* **2009**, *37*, 821–847. [CrossRef]
51. Gelbrich, K.; Hagel, J.; Orsingher, C. Emotional support from a digital assistant in technology-mediated services: Effects on customer satisfaction and behavioral persistence. *Int. J. Res. Mark.* **2021**, *38*, 176–193. [CrossRef]
52. Horning, M.A. Interacting with news: Exploring the effects of modality and perceived responsiveness and control on news source credibility and enjoyment among second screen viewers. *Comput. Hum. Behav.* **2017**, *73*, 273–283. [CrossRef]
53. Goodhue, D.; Thompson, R.L. Task-technology fit and individual performance. *MIS Q.* **1995**, *19*, 213–236. [CrossRef]

54. Zhou, T.; Lu, Y.; Wang, B. Integrating TTF and UTAUT to explain mobile banking user adoption. *Comput. Hum. Behav.* **2010**, *26*, 760–767. [CrossRef]
55. Ponsignon, F. Making the customer experience journey more hedonic in a traditionally utilitarian service context: A case study. *J. Serv. Manag.* **2023**, *34*, 294–315. [CrossRef]
56. Hill, J.; Randolph Ford, W.; Farreras, I.G. Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Comput. Hum. Behav.* **2015**, *49*, 245–250. [CrossRef]
57. Zhang, J.; Zhu, Y.; Wu, J.; Yu-Buck, G.F. A natural apology is sincere: Understanding chatbots' performance in symbolic recovery. *Int. J. Hosp. Manag.* **2023**, *108*, 103387. [CrossRef]
58. Hua, S.Y.; Jin, S.C.; Jiang, S.Y. The limitations and ethical considerations of chatgpt. *Data Intell.* **2024**, *6*, 201–239. [CrossRef]
59. Azaria, A.; Azoulay, R.; Reches, S. Chatgpt is a remarkable tool-for experts. *Data Intell.* **2024**, *6*, 240–296. [CrossRef]
60. Heather, M.; Gray, K.; Wegner, D.M. Dimensions of mind perception. *Science* **2007**, *315*, 619. [CrossRef]
61. Chen, G.; Dang, J.; Liu, L. After opening the black box: Meta-dehumanization matters in algorithm recommendation aversion. *Comput. Hum. Behav.* **2024**, *161*, 108411. [CrossRef]
62. Zhang, J.; Wang, X.; Lu, J.; Liu, L.; Feng, Y. The impact of emotional expression by artificial intelligence recommendation chatbots on perceived humanness and social interactivity. *Decis. Support Syst.* **2024**, *187*, 114347. [CrossRef]
63. Taheri, B.; Olya, H.; Ali, F.; Gannon, M.J. Understanding the influence of airport services cape on traveler dissatisfaction and misbehavior. *J. Travel Res.* **2020**, *59*, 1008–1028. [CrossRef]
64. Borah, S.B.; Prakhya, S.; Sharma, A. Leveraging service recovery strategies to reduce customer churn in an emerging market. *J. Acad. Mark. Sci.* **2020**, *48*, 848–868. [CrossRef]
65. Wang, Y.S.; Wu, S.C.; Lin, H.H.; Wang, Y.Y. The relationship of service failure severity, service recovery justice and perceived switching costs with customer loyalty in the context of e-tailing. *Int. J. Inf. Manag.* **2011**, *31*, 350–359. [CrossRef]
66. Schuetzler, R.M.; Grimes, G.M.; Giboney, J.S.; Rosser, H.K. Deciding whether and how to deploy chatbots. *MIS Q. Exec.* **2021**, *20*, 1–15. [CrossRef]
67. Chang, W.; Kim, K. Appropriate service robots in exchange and communal relationships. *J. Bus. Res.* **2022**, *141*, 462–474. [CrossRef]
68. Shan, M.; Zhu, Z.; Chen, H.; Sun, S. Service robot's responses in service recovery and service evaluation: The moderating role of robots' social perception. *J. Hosp. Mark. Manag.* **2023**, *33*, 145–168. [CrossRef]
69. Obeidat, Z.M.I.; Xiao, S.H.; Iyer, G.R.; Nicholson, M. Consumer revenge using the internet and social media: An examination of the role of service failure types and cognitive appraisal processes. *Psychol. Mark.* **2017**, *34*, 496–515. [CrossRef]
70. Sengupta, A.; Balaji, M.S.; Krishnan, B. How customers cope with service failure? A study of brand reputation and customer satisfaction. *J. Bus. Res.* **2015**, *68*, 665–674. [CrossRef]
71. Jerger, C.; Wirtz, J. Service employee responses to angry customer complaints: The roles of customer status and service climate. *J. Serv. Res.* **2017**, *20*, 362–378. [CrossRef]
72. Lv, X.; Liu, Y.; Luo, J.; Liu, Y.; Li, C. Does a cute artificial intelligence assistant soften the blow? The impact of cuteness on customer tolerance of assistant service failure. *Ann. Tour. Res.* **2021**, *87*, 103114. [CrossRef]
73. Wang, X.; Guchait, P.; Paşamehmetoğlu, A. Tolerating errors in hospitality organizations: Relationships with learning behavior, error reporting and service recovery performance. *Int. J. Contemp. Hosp. Manag.* **2020**, *32*, 2635–2655. [CrossRef]
74. Roehm, M.L.; Brady, M.K. Consumer responses to performance failures by high-equity brands. *J. Consum. Res.* **2007**, *34*, 537–545. [CrossRef]
75. Rapp, A.; Curti, L.; Boldi, A. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *Int. J. Hum.-Comput. Stud.* **2021**, *151*, 102630. [CrossRef]
76. Townsend, C.; Sood, S. Self-affirmation through the choice of highly aesthetic products. *J. Consum. Res.* **2012**, *39*, 415–428. [CrossRef]
77. Smith, A.K.; Bolton, R.N.; Wagner, J. A model of customer satisfaction with service encounters involving failure recovery. *J. Mark. Res.* **1999**, *36*, 356–372. [CrossRef]
78. Hayes, A. Introduction to mediation, moderation, and conditional process analysis. *J. Educ. Meas.* **2013**, *51*, 335–337.
79. Wang, C.; Li, Y.; Fu, W.; Jin, J. Whether to trust chatbots: Applying the event-related approach to understand consumers' emotional experiences in interactions with chatbots in e-commerce. *J. Retail. Consum. Serv.* **2023**, *73*, 103325. [CrossRef]
80. Yuan, C.; Zhang, C.; Wang, S. Social anxiety as a moderator in consumer willingness to accept AI assistants based on utilitarian and hedonic values. *J. Retail. Consum. Serv.* **2022**, *65*, 102878. [CrossRef]
81. Lin, H.; Chi, O.H.; Gursoy, D. Antecedents of customers' acceptance of artificially intelligent robotic device use in hospitality services. *J. Hosp. Mark. Manag.* **2019**, *29*, 530–549. [CrossRef]
82. Zhu, Y.; Zhang, J.; Wu, J.; Liu, Y. AI is better when I'm sure: The influence of certainty of needs on consumers' acceptance of AI chatbots. *J. Bus. Res.* **2022**, *150*, 642–652. [CrossRef]

83. Fan, A.; Wu, L.; Miao, L.; Mattila, A.S. When does technology anthropomorphism help alleviate customer dissatisfaction after a service failure?—The moderating role of consumer technology self-efficacy and interdependent self-construal. *J. Hosp. Mark. Manag.* **2020**, *29*, 269–290. [CrossRef]
84. Leo, X.; Huh, Y.E. Who gets the blame for service failures? Attribution of responsibility toward robot versus human service providers and service firms. *Comput. Hum. Behav.* **2020**, *113*, 106520. [CrossRef]
85. Wu, J.; Na, R.S.; Zheng, Y.X. Analysis of consumer complaints of the e-commerce to help farmers based on online reviews. *J. Syst. Eng.* **2025**, *40*, 15–32.
86. Su, Y.L.; Du, Z.Q. Data risk warning model of Internet entrepreneurship under the background of digitalization. *J. Syst. Eng.* **2025**, *40*, 215–231.
87. Fedorikhin, A.; Park, C.W.; Thomson, M. Beyond fit and attitude: The effect of emotional attachment on consumer responses to brand extensions. *J. Consum. Psychol.* **2008**, *18*, 281–291. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

systems

MDPI

Article

Shipping Market Sentiment Shocks and BDI Volatility: Evidence from News-Based Indicators

Lili Qu, Nan Hong * and Yutong Han

School of Maritime Economics and Management, Dalian Maritime University, Dalian 116026, China; qulili@dlmu.edu.cn (L.Q.); wutongyu5077@dlmu.edu.cn (Y.H.)
* Correspondence: woshihongnan@dlmu.edu.cn

Abstract

To address the lag and limited sensitivity of conventional shipping freight indicators, this study develops a news-based sentiment measure for the shipping market and examines its association with BDI dynamics. Using shipping-related news headlines from 2019 to 2025, a RoBERTa classifier fine-tuned on manually annotated data is used to quantify headline sentiment, and a daily Cumulative Sentiment Index (CSI) is constructed using an event-smoothing window with exponential decay. A higher CSI indicates more positive market sentiment, while a lower CSI reflects more negative sentiment. Empirical evidence shows that CSI exhibits pronounced responses around major market events and is closely linked to BDI behavior in event windows. In addition, an EGARCH specification augmented with CSI indicates that sentiment is significantly associated with conditional volatility, suggesting that news-based sentiment contains incremental information relevant to BDI risk dynamics. Overall, the proposed CSI provides a quantitative approach to tracking shipping market sentiment using publicly available news headlines and offers a complementary perspective to transaction-based freight indices.

Keywords: shipping industry; news sentiment; RoBERTa; EGARCH; risk warning

1. Introduction

The rising global economic uncertainty, coupled with frequent geopolitical conflicts and public events, has led to significant volatility in the shipping market, which serves as a core pillar of international trade. Traditional shipping freight indices, such as the Baltic Dry Index (BDI), while accurately measuring market conditions based on historical transaction data, suffer from substantial lag and may fail to reflect contemporaneous shifts in market sentiment, such as changes in macroeconomic expectations or perceived risks. For instance, during the initial outbreak of the Red Sea crisis in 2023, shipping companies, concerned about potential disruptions to trade routes, swiftly adjusted vessel deployment plans, and pessimistic expectations spread rapidly across the market. Yet the BDI did not exhibit noticeable changes until after the event had unfolded, limiting its usefulness for timely risk monitoring. This discrepancy highlights the need for complementary indicators that can track sentiment and risk perceptions more promptly.

In financial economics, a growing body of evidence suggests that sentiment indices derived from news can help explain market behavior and are closely related to volatility dynamics. Tetlock et al. [1], for example, show that news sentiment contains informative content for short-horizon market movements, offering a useful reference for shipping markets where expectations can adjust quickly in response to unfolding information.

Building on this line of research, this paper constructs a news-based sentiment indicator for the shipping market and examines its relationship with BDI behavior. To summarize time-varying sentiment signals in a tractable manner, sentiment scores are aggregated using a finite post-publication influence window and exponential decay, thereby allowing the impact of earlier information to diminish as new information arrives.

Specifically, sentiment signals are quantified from shipping-related news headlines using a RoBERTa-based classifier, and these signals are aggregated into a daily Cumulative Sentiment Index (CSI) via an event-smoothing structure with exponential decay. The empirical analysis focuses on two related questions: whether CSI exhibits systematic co-movement with BDI around major shipping-related events, and whether CSI is significantly associated with time-varying volatility dynamics beyond standard persistence and asymmetry effects. Accordingly, we formulate the following testable hypotheses:

H1: *Sentiment shocks captured by extreme movements in CSI are associated with systematic responses in BDI returns.*

H2: *Higher levels of CSI are positively associated with increased BDI volatility, indicating that sentiment fluctuations are linked to heightened market risk.*

To evaluate these hypotheses, we employ an event-study design to examine market reactions around sentiment shock events and an EGARCH-based volatility framework to assess the relationship between CSI and conditional volatility. Overall, the study contributes a news-headline-based sentiment indicator tailored to the shipping market and provides empirical evidence on its association with BDI movements and volatility dynamics, offering a quantitative complement to transaction-based freight indices.

2. Literature Review

2.1. Shipping Market Dynamics and Volatility Modeling

Existing research on shipping markets emphasizes freight rate fluctuations, nonlinear behavior, and volatility clustering, with particular attention to dry bulk shipping and the BDI as core indicators.

2.1.1. Research on the BDI and Shipping Market Dynamics

Shipping market dynamics have been widely studied with an emphasis on nonlinearity, structural change, and sensitivity to external shocks, with the BDI and dry bulk freight rates serving as core indicators. Ko [2] documents regime-dependent behavior in dry bulk freight markets, showing that volatility and adjustment mechanisms differ significantly across market states. Meng et al. [3] further reveal that shipping markets exhibit the pronounced nonlinear and asymmetric responses to geopolitical risks, particularly under extreme conditions. Riaz et al. [4] demonstrate that energy-related uncertainty generates significant and time-varying spillovers to maritime freight markets, with risk transmission intensifying during major economic and energy events. Feng and Chen [5] show that shipping market connectedness with green finance and climate risks is highly state- and frequency-dependent, becoming substantially stronger under extreme market scenarios. Chen et al. [6] provide evidence of asymmetric and time-varying cross-market risk contagion between fossil fuel and shipping markets, highlighting the role of external shocks in amplifying freight rate volatility. From a broader perspective, Mo et al. [7] review recent advances in shipping economic forecasting and note that, despite methodological progress, most studies remain heavily reliant on historical price-based variables, which may limit their effectiveness under abrupt structural changes.

https://doi.org/10.3390/systems14030317

2.1.2. Applications of EGARCH in Shipping Markets

Volatility modeling has long been a central topic in shipping market research, as freight rates and shipping-related assets exhibit pronounced volatility clustering and asymmetric responses to shocks. Early studies document strong dynamic interactions between freight volatility and market fundamentals, indicating that shipping markets are characterized by persistent and time-varying conditional heteroskedasticity [8]. Building on this foundation, GARCH-family models have been widely adopted to capture volatility dynamics in freight, shipping stock, and forward freight agreement markets, highlighting the importance of modeling conditional variance in shipping-related time series [9,10].

More recent research emphasizes the asymmetric and leverage effects inherent in shipping market volatility. Kim [11] explicitly applies an ARIMA–EGARCH framework to shipping-related indices and demonstrates that negative shocks exert a stronger impact on volatility than positive shocks, particularly during crisis periods. Gavriilidis et al. [12] further show that incorporating oil price shocks into GARCH-based models significantly improves volatility forecasts for tanker freight rates, reinforcing the relevance of external drivers in volatility formation. Extensions incorporating dynamic connectedness frameworks reveal that volatility spillovers between shipping and energy markets are time-varying and state-dependent, with risk transmission intensifying during turbulent periods [13]. Recent empirical evidence from tanker freight markets confirms that volatility dynamics remain highly persistent and asymmetric across shipping segments, underscoring the suitability of EGARCH-type specifications for capturing nonlinear and asymmetric risk characteristics in shipping markets [14].

2.2. Sentiment Analysis and Market Applications

Recent studies highlight the role of sentiment analysis in extracting information beyond traditional price-based indicators, especially in environments characterized by uncertainty and abrupt market changes.

2.2.1. Application of Sentiment Analysis in Market Studies

Early studies that extract sentiment from textual data commonly relied on dictionary-based approaches using predefined word lists. However, Loughran and McDonald [15] show that general-purpose sentiment dictionaries may produce substantial misclassification in financial texts due to the domain-specific language. Although domain-specific dictionaries may partially address this issue, their construction often requires substantial expert knowledge and data resources. With the development of natural language processing techniques, more studies have begun to employ machine learning-based methods for sentiment extraction.

Beyond measurement methods, a growing body of literature examines how sentiment indices influence financial market dynamics. Yousaf et al. [16] show that investor sentiment exhibits relatively weak average connectedness with financial markets but becomes a major transmitter of spillovers in the extreme upper and lower tails of return distributions, indicating that sentiment plays a critical role in periods of market stress. Focusing on global transmission mechanisms, Bei et al. [17] find that investor sentiment primarily drives short-term spillovers across financial markets and significantly magnifies the effects of capital flows and unconventional monetary policies during distress episodes, thereby accelerating the accumulation of systemic risk. Comparative evidence from the U.S. and Chinese stock markets further suggests that sentiment effects are highly dependent on institutional and behavioral characteristics, with stronger sensitivity and risk aversion observed in markets where information frictions and behavioral biases are more pronounced [18].

Beyond aggregate spillovers, sentiment has also been shown to influence expectation formation, volatility dynamics, and market predictability at the micro and applied levels. Frijns et al. [19] demonstrate that sentiment directly affects investors' expectations and increases heterogeneity across belief groups, reinforcing nonlinear adjustment processes in asset prices. In derivatives markets, Bahloul and Dammak [20] show that sentiment indices constructed using domain-specific lexicons significantly improve short-term return predictability, highlighting the importance of tailoring sentiment measurement to market-specific language and structures. Evidence from crisis periods further indicates that negative sentiment disproportionately amplifies volatility and downside risk, even in the presence of severe exogenous shocks such as the COVID-19 pandemic [21]. Related research on online collective action emphasizes how emotional contagion and opinion leadership can accelerate information diffusion and intensify collective responses [22], offering behavioral insights into how sentiment-driven dynamics may propagate in markets characterized by strong event dependence and rapid expectation shifts.

2.2.2. Research on the Shipping Market Sentiment

Existing studies on shipping sentiment mainly extract market emotions from shipping-related textual information and examine their role in explaining and predicting freight rate dynamics. Early research constructs news-based sentiment indices for dry bulk markets and finds that shipping sentiment closely tracks major freight indices and significantly predicts future freight rates, particularly in large vessel segments such as Capesize, while sentiment shocks tend to be transitory and asymmetric, with stronger effects under extreme market conditions [23]. More recent work applies market-specific language models to construct sentiment indices across dry bulk, tanker, and container markets, showing that sentiment serves as a Granger cause of key freight indices and that language-model-based measures substantially outperform traditional lexicon-based approaches [24]. Commodity-oriented studies further document nonlinear and regime-dependent relationships between shipping sentiment and freight rates, with sentiment capturing event-driven uncertainty and exhibiting stronger predictive power under optimistic or extreme market regimes [25]. Beyond bulk shipping, sentiment indicators have also been integrated into system dynamics frameworks for container markets, where freight rates are shown to be highly sensitive to the interaction between supply, demand, and market sentiment, especially during supply chain disruptions [26].

Overall, existing shipping market studies largely rely on freight rate indices and price-based indicators, typically incorporating supply–demand conditions, macroeconomic variables, and policy factors for market monitoring and forecasting. Despite increasing model sophistication, these approaches remain limited in capturing implicit information related to policy shocks, abrupt events, and market sentiment. In contrast, text-based sentiment analysis has been widely shown in financial market research to extract forward-looking signals and complement traditional quantitative models, particularly under nonlinear dynamics and extreme market conditions. However, its application in the shipping market remains relatively underdeveloped, with existing studies mostly focusing on low-frequency sentiment measures or linear frameworks and paying limited attention to high-timeliness sentiment signals and their role in volatility amplification and latent risk transmission. Furthermore, the use of advanced techniques such as pretrained language models for measuring shipping market sentiment has yet to be systematically explored.

https://doi.org/10.3390/systems14030317

3. Methodology

A sentiment-based framework is developed to quantify shipping news, construct a cumulative sentiment index, and examine its interaction with freight rate dynamics. The overall technical framework is illustrated in Figure 1.

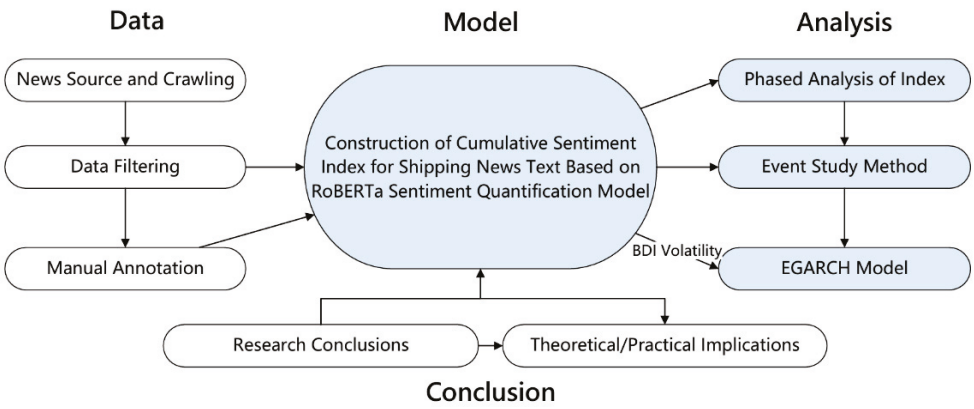


Figure 1. Technical Framework Diagram.

3.1. Theoretical Framework of Shipping News Sentiment

Shipping markets are highly sensitive to information related to global trade conditions, logistics disruptions, and geopolitical developments. As a widely disseminated information source, news not only reports objective events but also conveys interpretations of policy signals, market expectations, and perceived risks. In an industry characterized by long investment cycles, limited transparency, and strong exposure to external shocks, such information can influence market expectations before it is fully reflected in freight rate indicators such as the BDI. Consequently, shipping-related news provides an important channel through which market participants update expectations regarding freight demand, operational risks, and trade conditions.

In this study, sentiment is interpreted not merely as emotional tone but as a composite signal reflecting market expectations and risk perceptions embedded in news reports. According to media dependency theory [27], individuals and organizations rely more heavily on media information when environmental uncertainty and information asymmetry increase. This mechanism is particularly relevant in shipping markets, where freight rate dynamics are closely linked to macroeconomic conditions and geopolitical developments. Framing theory further suggests that news reporting is not neutral [28]; linguistic emphasis and narrative framing influence how market participants interpret events. Expressions such as "blockade," "disruption," or "supply shock" may therefore amplify perceived risk and alter expectations regarding shipping demand and operations. Empirical research also shows that media sentiment can influence financial market behavior by shaping investor expectations [1].

Sentiment signals may also diffuse across the market through behavioral mechanisms such as herding and information cascades [29]. When early responses by some market participants—often perceived as better informed—trigger observable operational or investment adjustments, other firms may imitate these actions even without complete information. Such imitation accelerates the diffusion of sentiment across heterogeneous market participants in the shipping ecosystem, strengthening collective responses and increasing the likelihood of abrupt adjustments in freight rates and volatility clustering. Similar diffusion mechanisms have also been identified in related economic contexts where information or value signals propagate across interconnected agents within supply chains [30], suggesting that sentiment signals may likewise spread among heterogeneous participants in the shipping ecosystem. Quantifying shipping news sentiment therefore provides a systematic way

to capture shifts in market expectations and risk perceptions that may precede observable movements in freight returns and volatility.

3.2. Data Selection and Preprocessing

3.2.1. Approach to the Selection of Shipping News Text Data

To ensure the scientific validity and representativeness of the sentiment index, the shipping news texts in this study are selected and defined based on three criteria: text boundaries, time frame, and analysis unit.

(1) Text Boundaries

As a highly globalized industry, the shipping sector generates public information primarily through specialized industry media and professional information platforms. Accordingly, this study collects shipping-related news headlines from authoritative sources, including China Shipping Network, China Waterway Network, Port Network, and Logistics Baba, which provide timely coverage of developments relevant to shipping market conditions. To ensure thematic relevance and semantic consistency, the data are filtered to retain only news directly or indirectly reflecting changes in shipping operations, market conditions, or related supply chain dynamics, while interviews, promotional content, and other non-market-oriented texts are excluded through a combination of keyword-based screening and manual verification.

(2) Time Frame

The study period is set from 2019 to early 2025. This time frame not only encompasses significant events that led to dramatic changes in global trade and shipping structures, such as the COVID-19 pandemic and the Red Sea crisis, but also limits the impact of changes in news websites' own development on the quantity of reports and narrative frameworks. By constructing a dataset of news headlines during this period, which includes multiple major market shocks, the sentiment index is better equipped to reflect the true evolution of shipping public opinion, thereby enhancing its timeliness and representativeness.

(3) Analysis Unit

Consistent with existing research in shipping sentiment analysis (e.g., Bai et al. [23]), news headlines are chosen as the unit of analysis in this study. On the one hand, as a highly condensed form of news content, headlines summarize key information and often reflect the main stance or emphasis of the news agency, thereby capturing the core sentiment conveyed by the event. On the other hand, headlines are concise and contain fewer irrelevant words than full news articles, which facilitates the extraction of salient textual features through natural language processing algorithms while reducing noise from lengthy contextual narratives. Although headlines may occasionally simplify complex events and introduce potential bias in sentiment interpretation, they remain an efficient textual unit for capturing timely market signals.

3.2.2. Manual Annotation Method for Sentiment Bias in News Headlines

Accurately annotating the sentiment bias in news headlines is crucial for training the sentiment quantification model. Given the abundance of industry-specific terminology and complex semantic structures in shipping news, general sentiment lexicons and unsupervised learning methods are often insufficient for effectively identifying the market signals embedded in the headlines. To address this, a rigorous manual annotation process is adopted to ensure the accuracy and industry consistency of the labels, providing high-quality training data for subsequent modeling.

(1) Annotation Targets and Classification System

This study focuses on the sentiment bias reflected in news headlines, categorizing them based on their potential impact on the shipping market. The dataset is divided into

three categories: (i) Positive: Reflects information that signals a positive market outlook, such as price increases, improved supply–demand conditions, smooth port operations, or favorable policies that boost industry confidence. (ii) Negative: Involves content that indicates negative market impacts, such as falling prices, port congestion, rising shipping costs, supply chain disruptions, or geopolitical risks. (iii) Neutral: Represents information with no clear market sentiment, such as neutral announcements, technical adjustments, or data releases. To clarify the annotation criteria, Table 1 provides several example headlines, their sentiment categories, and the rationale for each classification.

Table 1. Examples of Sentiment Category Determination and Rationale.

News Headline	Category	Rationale
"New Chongqing-to-Indonesia Iron-and-Sea Transport Line Launched"	Positive	The new transport line enhances logistics efficiency and promotes market diversification.
"Trump Appoints Panama Envoy, Takes Hardline Stance on Panama Canal"	Negative	The hardline policy may increase uncertainty in the shipping market.
"Concerns Over Strikes, Several Shipping Companies Announce U.S. East Coast Surcharge"	Negative	Concerns about strikes may lead to shipping disruptions, affecting the market.
"Lianyungang Port Group Plans to Transfer 18% Stake in Lianyungang Port Co. to Shanghai Port Group"	Neutral	The transfer may bring in liquidity but could affect the port's strategic control.

(2) Annotation Procedure

To ensure label consistency and accuracy, a team of three researchers with experience in shipping economics was organized for the annotation task. A "dual-annotator + one reviewer" mechanism was adopted. The specific process is as follows: (i) Pre-training and Annotation Guidelines: Before formal annotation, the team members were trained on shipping terminology and typical sentiment expression cases, and a detailed annotation manual was created. (ii) Initial Annotation: Two annotators independently labeled the same batch of headlines and recorded their sentiment labels. (iii) Consistency Check and Review: If the annotations were consistent, they were accepted as the final labels. In case of discrepancies, a third annotator made the final decision to ensure accurate sentiment reflection. (iv) Iterative Optimization: During the annotation process, ambiguous or difficult-to-classify samples were continuously reviewed and feedback was provided to refine the annotation guidelines and improve overall consistency. Random sampling was employed to ensure balanced coverage across time, platforms, and categories, thus avoiding selection bias.

(3) Quality Control and Consistency Evaluation

To assess the reliability of the annotation results, Cohen's Kappa coefficient is employed to evaluate inter-annotator agreement during the labeling process. In an initial pilot annotation of 200 samples, the agreement between two annotators reaches 0.84, indicating a high level of consistency in sentiment classification. As the annotation guidelines are refined and representative cases accumulate, the average Kappa value for the full dataset remains above 0.80, ensuring the stability and reusability of the sentiment labels.

3.3. Model and Index Construction Methods

3.3.1. Headline-Level Sentiment Scoring Model

RoBERTa is a Transformer-encoder language model. In this study, it serves as a contextual feature extractor for headline-level sentiment classification. Its core operation is

self-attention, which assigns context-dependent weights to tokens. Given query (Q), key (K), and value (V) projections, the scaled dot-product attention is defined as:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^{\mathsf{T}}}{\sqrt{d_k}}\right)V \tag{1}$$

RoBERTa is pre-trained with masked language modeling (MLM) using dynamic masking and does not include the next sentence prediction (NSP) task. For downstream sentiment classification, the final-layer representation of the [CLS] token is mapped to class probabilities via:

$$\hat{y} = Softmax(Wh_{[CLS]} + b) \tag{2}$$

where $\hat{y}$ denotes the predicted probability vector over negative, neutral, and positive classes, and $h_{[CLS]}$ is the final-layer embedding of the [CLS] token. A continuous headline-level sentiment score s_i is then obtained by a linear mapping of $\hat{y}$, which is subsequently used for CSI construction.

Specifically, the headline-level sentiment score s_i is defined through a linear polarity mapping that assigns weights $+1$, 0, and -1 to the positive, neutral, and negative classes, respectively:

$$s_i = (1)p_{pos} + (0)p_{neu} + (-1)p_{neg} \tag{3}$$

where p_{pos}, p_{neu}, and p_{neg} denote the predicted probabilities of the positive, neutral, and negative classes. This transformation yields a continuous sentiment measure in $[-1, 1]$, which is subsequently used as the headline-level input in the construction of CSI.

3.3.2. Construction of the Cumulative Sentiment Index

Market sentiment in the shipping industry is rarely driven by a single piece of information. Instead, it emerges from the accumulation of multiple news signals that gradually shape market expectations over time. To capture this dynamic information aggregation process, this study constructs the CSI that summarizes the sentiment signals contained in shipping-related news headlines.

Each news headline is evaluated using a trained sentiment classification model, which assigns a sentiment score s_i to headline i. Rather than treating each news item as an isolated observation, the CSI aggregates these sentiment signals across time while allowing their influence to decay as new information arrives. This structure reflects how market attention evolves in practice: recently released information tends to exert a stronger influence on market perceptions, whereas older information gradually loses relevance as subsequent developments emerge.

Formally, the CSI is defined as a time-aggregated sentiment measure in which the contribution of each news item is distributed across a finite influence period following its publication. The daily value of the index is calculated as:

$$CSI_d = \sum_{i:t_i \le d \le t_i + W} s_i \cdot e^{-\lambda(d - t_i)} \tag{4}$$

where d denotes the target date, s_i represents the sentiment score assigned to news item i, and t_i is the publication date of that item. The parameter W specifies the length of the influence window during which a news item contributes to the index, while λ controls the rate at which the influence of past information decays over time.

Parameter values are calibrated to reflect the characteristics of the shipping market. The influence window W of each news item is restricted to 7 consecutive days following publication, reflecting the empirical observation that news-driven sentiment effects in

https://doi.org/10.3390/systems14030317

shipping markets are typically most pronounced within one week and diminish thereafter as market attention wanes.

The decay parameter is set to λ = 0.1, under which sentiment influence declines to approximately one-third of its initial magnitude after ten days. During model calibration, alternative specifications of both the event window length and the decay parameter were evaluated. The forecasting performance remains quantitatively similar across reasonable parameter choices, and the main empirical conclusions are unaffected. We therefore adopt this specification as a parsimonious and economically consistent benchmark.

3.3.3. EGARCH Model Mechanism and Extensions

In both shipping and financial markets, price volatility often exhibits two well-documented stylized facts: volatility clustering and leverage effects. Volatility clustering refers to the tendency for periods of high volatility to be followed by further high volatility, while tranquil periods are followed by relatively small fluctuations. Leverage effects describe the asymmetric response of volatility to shocks, where negative news tends to generate a larger increase in market volatility than positive news of the same magnitude. To capture these empirical features in the BDI market, an Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) framework is adopted.

In an EGARCH model, the dynamic evolution of volatility is characterized by a logarithmic conditional variance equation. Specifically, the EGARCH(1,1) specification is written as:

$$\ln\left(\sigma_t^2\right) = \omega + \alpha\left(\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - E|z|\right) + \gamma\frac{\epsilon_{t-1}}{\sigma_{t-1}} + \beta\ln\left(\sigma_{t-1}^2\right) \tag{5}$$

where σ_t^2 denotes the conditional variance, β measures volatility persistence, α captures the magnitude effect of standardized shocks, and γ reflects asymmetry in the response of volatility to positive and negative innovations. A negative γ indicates the presence of a leverage effect. The centering term $E|z|$ is distribution-dependent and ensures that the shock component has zero mean under the assumed innovation distribution.

To examine whether shipping news sentiment contributes to volatility formation, the variance equation is augmented by incorporating the sentiment index:

$$\ln(\sigma_t^2) = \omega + \alpha\left(\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - E|z|\right) + \gamma\frac{\epsilon_{t-1}}{\sigma_{t-1}} + \beta\ln\left(\sigma_{t-1}^2\right) + \delta S_t^{(w)} \tag{6}$$

where $S_t^{(w)}$ denotes the weighted lagged sentiment index constructed from the cumulative sentiment series. The lag structure is determined based on the results of the Granger causality tests, which identify the statistically significant transmission horizon of sentiment to market volatility. The weighted combination therefore reflects the lag periods exhibiting significant predictive effects, allowing for delayed information diffusion and gradual market adjustment. The coefficient δ measures the marginal effect of sentiment on conditional volatility. A significantly positive δ indicates that stronger sentiment signals are associated with higher conditional variance, implying that news-based sentiment contributes to market risk dynamics. The model is estimated under a skewed Student-t distribution to accommodate excess kurtosis and asymmetry commonly observed in shipping market return series.

4. Sentiment Model and Index Construction

Shipping news headlines and the BDI data are processed to construct the sentiment measurement framework. A RoBERTa-based classifier is employed to extract sentiment signals from news headlines, and these signals are aggregated into the CSI that captures the temporal evolution of shipping market sentiment.

4.1. Data Preprocessing

This study combines shipping news headlines with freight market data to investigate the relationship between media sentiment and shipping market dynamics. Two types of data are used in the empirical analysis: shipping news headlines and the BDI.

The BDI serves as a widely recognized benchmark for conditions in the global dry bulk shipping market. Compiled and published by the Baltic Exchange in London, the index reflects daily freight rate movements for major bulk commodities such as iron ore, coal, and grain, which are closely linked to global trade activity and industrial demand. The sample period spans from November 2019 to January 2025, covering several major episodes of market volatility, including the COVID-19 pandemic, the Suez Canal blockage, and disruptions in the Red Sea shipping routes. Since the BDI is released only on working days, linear interpolation is applied to construct a continuous daily time series. After aligning the interpolated BDI series with the news dataset, the final sample consists of 1888 daily observations.

Shipping news headlines are collected from several major maritime information platforms and financial news sources that regularly report developments in global shipping markets. After removing duplicate records and filtering irrelevant content, a total of 20,656 shipping-related news headlines are retained. To preserve the temporal structure of the dataset and avoid information leakage, the headline data used for sentiment modeling are partitioned strictly according to chronological order. Specifically, the dataset is divided into training and testing subsets using an 80–20 split, where earlier observations are used for model training while the most recent observations are reserved for testing. This time-ordered partition ensures that the evaluation results are based on out-of-sample predictions, providing a realistic assessment of the model's ability to extract sentiment signals from newly arriving news.

4.2. Model Specification and Evaluation

The RoBERTa tokenizer is used to perform subword tokenization on the news headlines, with a maximum sequence length of 128 tokens. Sequences exceeding this length are truncated, while shorter sequences are padded with the [PAD] token to maintain consistent batch input. For model construction, the RoBERTa-based-Chinese pre-trained model from HuggingFace's Transformers library is adopted. Two fully connected layers together with a dropout layer are added to the output layer for the classification task to mitigate overfitting. The model is trained using the AdamW optimizer and cross-entropy loss, with early stopping and learning-rate scheduling applied to improve training stability.

The classification results indicate that 52.08% of the headlines are predicted as positive sentiment, 45.88% as negative sentiment, and 2.04% as neutral sentiment. The relatively small proportion of neutral labels reflects the nature of shipping news reporting, which often focuses on developments with clear market implications, such as freight rate changes, supply disruptions, and geopolitical events affecting shipping routes. Given this distribution, model performance is evaluated using Weighted-F1, Macro-F1, and Accuracy, which provide complementary measures of classification performance under potential class imbalance. All evaluation metrics are computed on the out-of-sample test set, ensuring that the reported results reflect the model's ability to generalize to newly arriving news data. The comparative performance of the models is reported in Table 2.

As reported in Table 2, the benchmark models achieve progressively improved performance in the sentiment classification task. Traditional text representation approaches such as Word2Vec-MLP and FastText provide a basic baseline for sentiment identification, while sequence-based architectures including TextCNN and BiLSTM improve performance by capturing contextual and sequential patterns in the news headlines. Pretrained language

models further enhance the extraction of semantic information from shipping-related texts. Among the evaluated models, RoBERTa achieves the highest scores across all evaluation metrics. This result suggests that contextual language representations are particularly effective for identifying sentiment signals in shipping news, which subsequently serve as inputs for the construction of the sentiment index.

Table 2. Performance comparison of sentiment quantification models.

Model	Train Loss	Weighted-F1	Macro-F1	Accuracy
Word2Vec_MLP	0.3151	0.8821	0.8500	0.8811
FastText	0.3364	0.8811	0.8458	0.8825
TextCNN	0.2018	0.9320	0.9137	0.9312
BiLSTM	0.2651	0.9256	0.9062	0.9244
BERT	0.0764	0.9669	0.9578	0.9666
RoBERTa	0.0672	0.9696	0.9615	0.9693

4.3. Cumulative Sentiment Index Analysis

The CSI constructed from shipping news headlines is illustrated in Figure 2, which depicts the evolution of market sentiment over the sample period. The index exhibits noticeable fluctuations around major developments in the shipping market, indicating that shipping-related news sentiment responds sensitively to changes in the global trade and transportation environment.

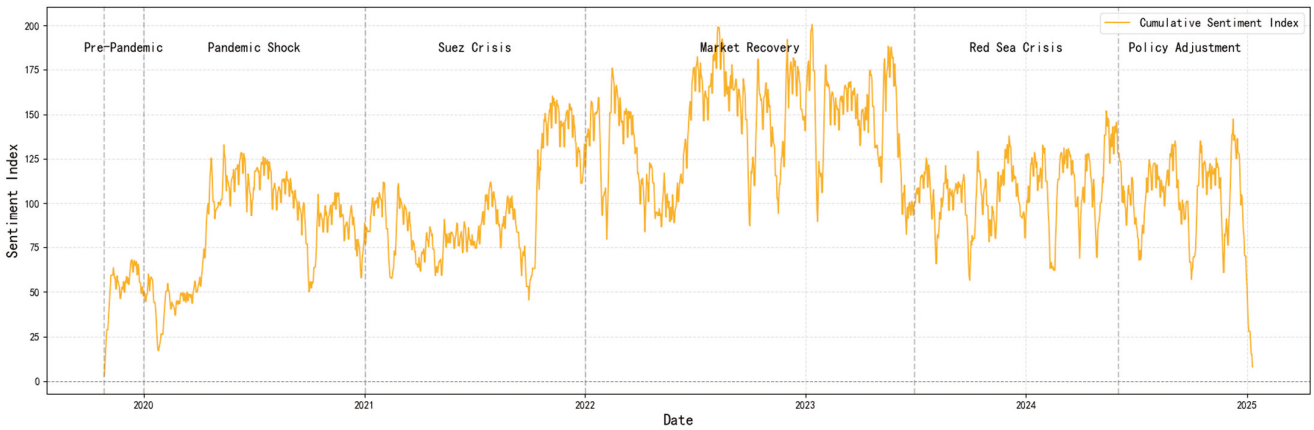


Figure 2. Shipping news cumulative sentiment index trend.

To further examine the statistical characteristics of the sentiment index, descriptive statistics are calculated across several periods defined according to major developments in the shipping market. These periods broadly correspond to key stages including the pre-pandemic period, the COVID-19 disruption phase, the Suez Canal blockage episode, the post-pandemic market recovery, and the Red Sea shipping crisis. The descriptive statistics of the CSI across these phases are reported in Table 3.

Table 3 reports the descriptive statistics of the CSI across different market phases. The results reveal clear stage-dependent variations in shipping-related news sentiment. During the Pre-COVID-19 period, the CSI remains at relatively low levels with limited variability, reflecting a comparatively stable market environment. Following the outbreak of COVID-19, the average CSI increases substantially, indicating intensified media attention and stronger sentiment responses to disruptions in global trade and logistics networks. Sentiment levels continue to rise during the Suez Canal shock and reach their highest

values during the post-pandemic recovery phase, consistent with the rebound in global freight demand and increased market activity.

Table 3. Descriptive statistics of the cumulative sentiment index across market phases.

Period	Observations	Mean	Std. Dev.	Skewness	Kurtosis
Pre-COVID-19 Period	65	52.4648	13.4933	−1.873	3.7738
COVID-19 Disruption	367	84.6621	28.7797	−0.4379	−1.0383
Suez Canal Shock	365	95.0151	27.3886	0.7995	−0.2748
Post-Pandemic Recovery	545	142.852	28.3082	−0.3409	−0.8991
Red Sea Crisis	337	107.1063	18.6653	−0.3656	0.0196
Recent Adjustment	223	101.3597	25.2542	−1.2218	1.8532

In subsequent periods, the CSI remains elevated but exhibits noticeable adjustments as new geopolitical and economic developments emerge. The Red Sea crisis is associated with persistently high sentiment levels, reflecting heightened market sensitivity to maritime security risks and shipping route disruptions. In the most recent adjustment phase, the CSI shows a moderate decline together with reduced dispersion, suggesting that market sentiment has gradually stabilized as the shipping industry adapts to evolving trade conditions. Overall, these descriptive patterns indicate that shipping news sentiment responds systematically to major events affecting global shipping markets, supporting the use of the CSI in the subsequent empirical analysis.

5. Results

The empirical analysis tests the relationship between sentiment shocks and BDI dynamics. Event study methods and an EGARCH framework are used to evaluate the effects of sentiment on freight returns and volatility.

5.1. Event Study of Sentiment Shocks and BDI Responses

5.1.1. Event Identification and Empirical Design

To further investigate the dynamic relationship between the CSI and the BDI, this study adopts an event study approach to quantify the short-term impact of sentiment shocks on the shipping market. Following the general framework of Tetlock et al. [1], sentiment shock events are identified based on the distribution of the CSI.

In terms of event definition, the daily values of the CSI are used to identify sentiment shocks. Following the standard quantile-based approach in the event study literature, a positive sentiment shock is defined as a trading day when the CSI exceeds the upper q-th quantile of its empirical distribution, while a negative sentiment shock corresponds to trading days when the CSI falls below the lower $(1-q)$-th quantile. According to this classification rule, the sample contains 112 positive sentiment shock events and 110 negative sentiment shock events.

For the event window design, considering that the shipping market may exhibit delayed responses to sentiment shocks, symmetric event windows are constructed around the event day. Specifically, event windows of ± 7, ± 14, and ± 30 days are considered in order to capture short-term and medium-term market reactions to sentiment shocks.

The daily logarithmic return of the BDI is used as the return measure. To evaluate the market response to sentiment shocks, the cumulative return (CR) within the event window

is calculated by aggregating the daily logarithmic returns around the event day. Formally, the cumulative return over the event window (τ_1, τ_2) is defined as

$$CR(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} R_t \tag{7}$$

where R_t denotes the logarithmic return of the BDI on day t. This cumulative measure captures the overall adjustment of BDI returns surrounding the identified sentiment shock events.

5.1.2. Event Study Results and Economic Interpretation

The cumulative return paths of BDI returns are examined around the identified sentiment shock days to characterize the market response within the event windows. Under the baseline threshold ($q = 90\%$), the evolution of cumulative returns is reported in Figure 3 for different event windows.

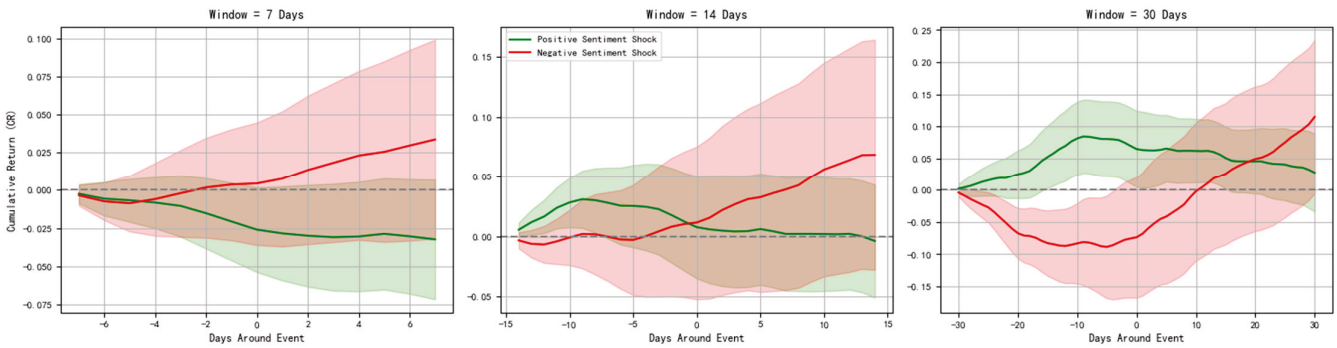


Figure 3. Event study results under the baseline threshold ($q = 90\%$).

The results show systematic differences between positive and negative sentiment shocks: positive shocks are associated with relatively mild movements, whereas negative shocks are followed by larger and more persistent declines in cumulative returns. This divergence becomes more pronounced as the event window expands, suggesting that sentiment effects may accumulate over time. Overall, the baseline event-study evidence supports H1 by showing that extreme sentiment movements are accompanied by systematic adjustments in BDI returns.

As a robustness check, sentiment shocks are re-identified using alternative quantile thresholds (q = 85%, 80%, and 75%) while keeping the event window fixed at ±14 days. The corresponding CR dynamics are presented in Figure 4. The results show that the overall response patterns remain consistent across different threshold definitions. In particular, negative sentiment shocks continue to generate stronger and more persistent market responses than positive shocks, and the divergence between the two responses becomes more evident after the event date. Although the magnitude of the responses varies slightly across thresholds, the overall shape and direction of the CR dynamics remain stable. These findings indicate that the observed return–response patterns are robust to alternative definitions of sentiment shock events.

From a practical perspective, the event-study results indicate that news-based sentiment signals can complement traditional freight indices in monitoring short-term freight market conditions. In particular, the pronounced response to negative sentiment shocks implies that adverse news coverage is associated with rapid adjustments in market expectations regarding freight demand and shipping activity. This pattern is consistent with the information-sensitive nature of shipping markets, where expectations about trade condi-

tions, logistics disruptions, and macroeconomic developments are gradually incorporated into freight rate dynamics.

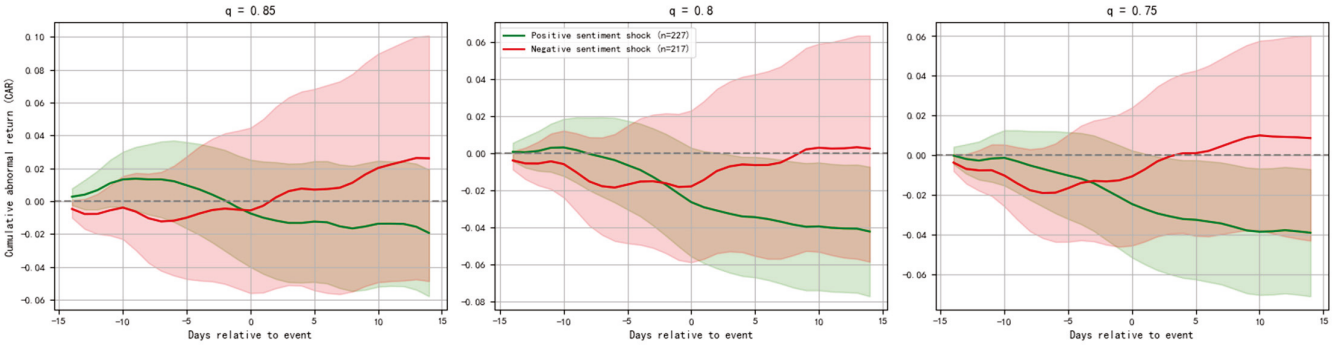


Figure 4. Robustness test using alternative sentiment thresholds (q = 85%, 80%, and 75%) under a fixed ±14-day event window.

Moreover, the asymmetric response between positive and negative sentiment shocks highlights the importance of monitoring downside information in freight markets. Negative news appears to transmit more strongly into market adjustments, suggesting that market participants may react more sensitively to signals of potential demand weakness or operational disruptions. Incorporating sentiment-based indicators into analytical frameworks may therefore help analysts and industry participants better interpret market signals and identify periods of heightened uncertainty. In this sense, sentiment measures can serve as a useful supplementary tool for understanding short-term fluctuations and expectation dynamics in global shipping markets.

5.2. EGARCH Analysis of the Sentiment Index and BDI Volatility

5.2.1. EGARCH Model Specification and Estimation

The EGARCH(1,1) model is estimated using the daily log return of the BDI as the dependent variable. Market sentiment is measured by the CSI. To allow for delayed sentiment transmission from news to market risk, the conditional variance equation is augmented with a lagged sentiment regressor $S_t^{(w)}$, constructed from past CSI values.

The lag structure is informed by Granger causality tests between CSI and BDI returns. The results show no statistically significant association at lag 1, whereas statistically significant lead–lag relationships emerge from lag 2 onward and remain pronounced over medium and longer horizons, with sustained significance observed particularly within the 16–28 lag range. To summarize these transmission horizons parsimoniously while avoiding an excessive number of lag terms, three representative significant lags (17, 20, and 28) are selected to construct $S_t^{(w)}$. A simple decreasing weighting scheme is adopted to place greater emphasis on the horizon where the transmission is most economically relevant while retaining information from adjacent significant lags. Specifically,

$$S_t^{(w)} = 0.6CSI_{t-17} + 0.3\ CSI_{t-20} + 0.1\ CSI_{t-28} \tag{8}$$

Parameters are estimated by maximum likelihood under a skewed Student-t innovation distribution, which is adopted to accommodate the skewness and heavy tails in BDI returns and provides the best overall fit among alternative distributions considered. All variables are aligned by trading dates, and observations with missing values are excluded to ensure a consistent estimation sample. Standard post-estimation diagnostics indicate an adequate fit. The ARCH-LM test on the residuals of the extended model yields a statistic of 12.36 (p = 0.196), suggesting no remaining conditional heteroskedasticity. Residual

autocorrelations are not statistically significant across conventional lag orders based on ACF/PACF checks, indicating that the standardized residuals are approximately serially uncorrelated. The corresponding estimation results are reported in Table 4 for specifications with and without the sentiment term.

Table 4. Comparison of EGARCH(1,1)-SKEWT model parameter estimation results.

Parameter	EGARCH with Sentiment	*p*-Value	EGARCH Without Sentiment	*p*-Value
Const	0.000265	0.0015	0.000270	0.0013
AR(1)	0.4881	0.0000	0.4926	0.0000
ω (omega)	−0.9688	0.0281	−0.9504	0.0301
δ (Sentiment)	0.1455	0.0216	-	-
α (alpha)	0.4482	0.0095	0.4555	0.0082
γ (gamma)	−0.1922	0.0103	−0.1844	0.0144
β (beta)	0.8623	0.0000	0.8587	0.0000
η (eta)	2.0481	0.0000	2.0500	0.0000
λ (lambda)	0.0398	0.0026	0.0395	0.0027

From the benchmark specification, the estimated volatility dynamics exhibit the standard stylized facts commonly observed in market data. The persistence parameter β is highly significant, indicating strong volatility persistence in BDI returns. The magnitude parameter α is significantly positive, suggesting that conditional volatility responds to the size of past shocks. The asymmetry parameter γ is significantly negative, implying that negative shocks generate a larger increase in volatility than positive shocks of the same magnitude, consistent with a leverage-type effect in the shipping market. The innovation distribution parameters η and λ are also statistically significant, supporting heavy tails and skewness in the return innovations and confirming that the skewed Student-t specification is appropriate.

After incorporating the CSI-based sentiment regressor, the extended model shows an improved overall fit based on information criteria, and the sentiment coefficient δ is positive and statistically significant at the 5% level. In economic terms, an increase in CSI is associated with a rise of 0.1455 in the log conditional variance of BDI returns, corresponding to an approximate 15.68% increase in conditional volatility. This provides support for H2 that higher CSI is associated with higher conditional volatility in the BDI. Substantively, the result indicates that news-based sentiment is closely linked to the time variation in shipping market risk: when sentiment signals intensify, the conditional volatility of freight-rate returns tends to be higher, reflecting a market environment characterized by greater uncertainty and more pronounced risk re-assessment.

5.2.2. Robustness Checks

To examine whether the estimated relationship between shipping sentiment and market volatility is sensitive to alternative specifications, two robustness checks are conducted.

First, to address the concern that the sentiment index may partly capture variations in news intensity rather than sentiment orientation itself, a news volume control variable V_t is introduced into the variance equation. The variable measures the daily number of shipping-related headlines and serves as a proxy for information arrival intensity. The

variance equation is therefore extended by including both the sentiment index $S_t^{(w)}$ and the news volume variable V_t:

$$\ln(\sigma_t^2) = \omega + \alpha\left(\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - E|z|\right) + \gamma\frac{\epsilon_{t-1}}{\sigma_{t-1}} + \beta\ln\left(\sigma_{t-1}^2\right) + \delta S_t^{(w)} + \theta V_t \tag{9}$$

Second, robustness is examined with respect to the lag specification of the sentiment variable. Instead of the weighted lag combination used in the baseline specification, alternative lag structures are considered, including a short-horizon average of recent CSI values and a single representative lag. The alternative sentiment variables are defined as:

$$S_t^{(1)} = \frac{1}{5}\sum_{i=1}^{5} CSI_{t-i} \tag{10}$$

$$S_t^{(2)} = CSI_{t-k} \tag{11}$$

The estimation results for the sentiment coefficient under alternative specifications are summarized in Table 5.

Table 5. Robustness checks for the sentiment effect in the EGARCH model.

Specification	δ (Sentiment)	*p*-Value	AIC	BIC
Baseline model	0.1455	0.0216	4448.64	4487.41
+ News volume control	0.1391	0.0289	4440.12	4486.75
$S_t^{(1)}$ (short-horizon mean)	0.1520	0.0176	4456.83	4503.46
$S_t^{(2)}$ (single lag)	0.1425	0.0248	4453.27	4499.9

As reported in Table 5, the sentiment coefficient remains positive and statistically significant across all specifications, with magnitudes broadly comparable to that of the baseline model. When the news volume variable is included, the sentiment coefficient decreases slightly but remains statistically significant, indicating that the sentiment index retains explanatory power beyond variations in news intensity. The moderate improvement in AIC after controlling for news volume does not materially alter the main findings. Results under alternative lag specifications are also consistent with the baseline model, confirming that the relationship between shipping sentiment and BDI volatility is robust.

5.2.3. Economic Interpretation

The empirical results indicate that CSI is positively associated with conditional volatility in BDI returns, suggesting that sentiment extracted from shipping-related news reflects fluctuations in market risk perceptions. In the shipping market, information related to trade demand, route stability, or logistical disruptions is rapidly incorporated into expectations about freight rates. News coverage therefore serves as an important channel through which market participants update their assessments of freight market risk.

One possible explanation is that shifts in sentiment increase heterogeneity in market expectations. When news signals convey stronger optimism or pessimism regarding shipping conditions, market participants may revise chartering strategies, route allocation decisions, and freight rate expectations. While such revisions do not necessarily change the average level of freight rates immediately, they can increase the dispersion of expectations, thereby intensifying short-term volatility as new information is gradually incorporated into prices.

The persistence and asymmetry observed in the volatility process further reflect structural characteristics of the shipping market. Negative shocks, such as geopolitical tensions, trade disruptions, or logistical bottlenecks, tend to generate stronger volatility responses than positive developments. This pattern is consistent with the asymmetric volatility effect captured by the EGARCH specification. Overall, these findings suggest that news-based sentiment indicators contain economically meaningful information about fluctuations in shipping market risk and can provide complementary signals for monitoring changes in freight market conditions alongside traditional price-based indicators.

6. Conclusions

This study develops a news-based sentiment indicator for the shipping market by combining manually annotated Chinese shipping news headlines with a RoBERTa-based sentiment classification model. By aggregating headline-level sentiment through an event-smoothing framework with exponential decay, the Cumulative Sentiment Index (CSI) captures the dynamic evolution of market expectations reflected in shipping-related news. Empirical results support the proposed hypotheses. The event-study analysis shows that extreme sentiment shocks are associated with systematic adjustments in BDI returns, with negative sentiment shocks producing stronger and more persistent responses than positive ones. In addition, the EGARCH estimation indicates that the CSI is positively and significantly associated with the conditional volatility of BDI returns, suggesting that sentiment signals embedded in news coverage contain incremental information about fluctuations in shipping market risk.

While the findings provide useful insights, several aspects of the study could be further improved. The sentiment index is constructed from shipping news headlines collected from a limited set of industry platforms, which may not fully represent the broader information environment of the shipping market; future research could therefore expand the data sources to include a wider range of media outlets and related domains to enhance the comprehensiveness of sentiment measurement; moreover, the empirical analysis mainly focuses on identifying the relationship between sentiment shocks and market dynamics rather than evaluating predictive performance. Future studies could incorporate forecasting frameworks to further assess the predictive value of news-based sentiment indicators. Overall, the results suggest that news-based sentiment measures can provide a useful complementary perspective for understanding information-driven fluctuations in the shipping market.

Author Contributions: Conceptualization, L.Q.; methodology, L.Q. and N.H.; software, L.Q. and N.H.; validation, L.Q.; data preprocessing and manual annotation, L.Q., N.H. and Y.H.; data curation, L.Q., N.H. and Y.H.; writing—original draft preparation, N.H. and Y.H.; writing—review and editing, L.Q.; visualization, N.H. and Y.H.; supervision, L.Q.; project administration, L.Q.; funding acquisition, L.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Social Science Fund of China, grant number 25BGJ073.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tetlock, P.C. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *J. Financ.* **2007**, *62*, 1139–1168. [CrossRef]
2. Ko, B.-W. A mixed-regime model for dry bulk freight market. *Asian J. Shipp. Logist.* **2010**, *26*, 185–204. [CrossRef]
3. Meng, B.; Chen, S.; Dong, J. Nonlinear risk dynamics and strategic management: Insights for LNG shipping policy and practice. *Transp. Policy* **2025**, *173*, 103752. [CrossRef]
4. Riaz, A.; Ahmed, S.; Nilu, T.Y.; Shahbaz, M. Spillover dynamics between maritime energy transportation and energy-related uncertainty. *Reg. Stud. Mar. Sci.* **2026**, *93*, 104703. [CrossRef]
5. Feng, A.; Chen, Y. Quantile time–frequency connectedness between green finance and shipping markets: The role of climate transition and physical risks. *Int. Rev. Econ. Financ.* **2025**, *104*, 104704. [CrossRef]
6. Chen, S.; Meng, B.; Lv, L. Cross-market risk contagion and portfolio diversification: A novel dynamic connectedness framework for fossil fuel and shipping markets. *Transp. Res. Part E* **2025**, *204*, 104443. [CrossRef]
7. Mo, J.; Gao, R.; Yuen, K.F.; Suganthan, P.N. Shipping economic forecasting: Recent developments, applications, and future directions. *Transp. Rev.* **2025**, *45*, 897–923. [CrossRef]
8. Xu, J.J.; Yip, T.L.; Marlow, P.B. The dynamics between freight volatility and fleet size growth in dry bulk shipping markets. *Transp. Res. Part E* **2011**, *47*, 983–991. [CrossRef]
9. Ruslan, S.M.M.; Mokhtar, K. Stock market volatility on shipping stock prices: GARCH models approach. *J. Econ. Asymmetries* **2021**, *24*, e00232. [CrossRef]
10. Alizadeh, A.H. Trading volume and volatility in the shipping forward freight market. *Transp. Res. Part E* **2013**, *49*, 250–265. [CrossRef]
11. Kim, C.B. Leverage effect of HRCI volatility and the volatility impact on Korean export container volume before and after the global financial crisis: Application of ARIMA–EGARCH and GIRF. *Asian J. Shipp. Logist.* **2018**, *34*, 227–233. [CrossRef]
12. Gavriilidis, K.; Kambouroudis, D.S.; Tsakou, K.; Tsouknidis, D.A. Volatility forecasting across tanker freight rates: The role of oil price shocks. *Transp. Res. Part E* **2018**, *118*, 376–391. [CrossRef]
13. Maitra, D.; Chandra, S.; Dash, S.R. Liner shipping industry and oil price volatility: Dynamic connectedness and portfolio diversification. *Transp. Res. Part E* **2020**, *138*, 101962. [CrossRef]
14. Ajith, P.J.; Raju, T.B.; Gupta, R.; Kulshrestha, N. Volatility in tanker freight markets. *Case Stud. Transp. Policy* **2023**, *12*, 100993. [CrossRef]
15. Loughran, T.; McDonald, B. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *J. Financ.* **2011**, *66*, 35–65. [CrossRef]
16. Yousaf, I.; Youssef, M.; Goodell, J.W. Quantile connectedness between sentiment and financial markets: Evidence from the S&P 500 Twitter sentiment index. *Int. Rev. Financ. Anal.* **2022**, *83*, 102322. [CrossRef]
17. Bei, Z.; Cui, L.; Zhou, Y. Liquidity, sentiment, and global spillover across financial markets. *J. Int. Money Financ.* **2026**, *161*, 103494. [CrossRef]
18. Zhang, Y.; Li, Y.; Wang, W. Comparative analysis of investor sentiment and attention in the U.S. and Chinese stock markets: Implications for financial stability. *Int. Rev. Financ. Anal.* **2026**, *109*, 104751. [CrossRef]
19. Frijns, B.; Huynh, T.; Zwinkels, R.C.J. Expectation formation in financial markets: Heterogeneity and sentiment. *J. Econ. Dyn. Control* **2025**, *177*, 105133. [CrossRef]
20. Bahloul, W.; Dammak, W. Online investor sentiment in the financial futures markets. *Res. Int. Bus. Financ.* **2025**, *75*, 102708. [CrossRef]
21. Bai, C.; Duan, Y.; Fan, X.; Tang, S. Financial market sentiment and stock return during the COVID-19 pandemic. *Financ. Res. Lett.* **2023**, *54*, 103709. [CrossRef]
22. Li, S.; Hao, J.; Chai, Y.; Liu, Y.; Feng, T.; Zhu, X. Emotional contagion mechanism and response strategy simulation of corporate crises from the perspective of online collective action. *Systems* **2026**, *14*, 1. [CrossRef]
23. Bai, X.; Lam, J.S.L.; Jakher, A. Shipping sentiment and the dry bulk shipping freight market: New evidence from newspaper coverage. *Transp. Res. Part E* **2021**, *155*, 102490. [CrossRef]
24. Sui, C.; Wang, S.; Zheng, W. Sentiment as a shipping market predictor: Testing market-specific language models. *Transp. Res. Part E* **2024**, *189*, 103651. [CrossRef]
25. Gong, Y.; Peng, Y.; Xu, L.; Chen, K.; Shi, W. Shipping news sentiment as a predictor of iron ore freight rates: Hybrid evidence from lexicon-based analysis and threshold autoregression modelling. *Transp. Policy* **2025**, *169*, 178–190. [CrossRef]
26. Jeon, J.; Iris, Ç.; Hong, S.; Lyons, A. Box rates unveiled: Predictive analytics for ocean freight rates with system dynamics and text mining under supply chain disruptions. *Int. J. Prod. Econ.* **2025**, *286*, 109669. [CrossRef]
27. Ball-Rokeach, S.J.; DeFleur, M.L. A dependency model of mass-media effects. *Commun. Res.* **1976**, *3*, 3–21. [CrossRef]
28. Entman, R.M. Framing: Toward clarification of a fractured paradigm. *J. Commun.* **1993**, *43*, 51–58. [CrossRef]

29. Bikhchandani, S.; Hirshleifer, D.; Welch, I. A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Polit. Econ.* **1992**, *100*, 992–1026. [CrossRef]
30. He, J.; Yang, Q.; Zhao, Y. Research on the diffusion effects of data value in supply chains. *J. Asian Econ.* **2025**, *101*, 102064. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

systems

MDPI

Article

Evolutionary Patterns and Advanced Strategies of Health Policies Based on Topic Modeling and Social Network Analysis

Kaixuan Zhu [1,†], Lirong Song [1], Xuejie Yang [2,*], Wenxing Lu [1,†] and Dongxiao Gu [1,3,*]

[1] School of Management, Hefei University of Technology, Hefei 230009, China; kaixuan_zhu@mail.hfut.edu.cn (K.Z.); luwenxing@hfut.edu.cn (W.L.)
[2] School of Humanity and Law, Hefei University of Technology, Hefei 230009, China
[3] Laboratory of Data Science and Smart Society Governance of the Ministry of Education, Hefei University of Technology, Hefei 230009, China
* Correspondence: xuejie_y@hfut.edu.cn (X.Y.); dongxiaogu@yeah.net (D.G.)
† These authors contributed equally to this work.

Abstract

We systematically analyze the evolutionary characteristics of China's public health policies, focusing on the dynamic changes in policy content, stage-specific differences, and inter-subject collaborative relationships. Based on 137 public health policy documents issued by the central government, the analysis is conducted from a dual perspective: first, the BERTopic model is employed to identify prominent policy themes and track their evolutionary paths; second, Social Network Analysis (SNA) is utilized to deconstruct the collaborative mechanisms and network structural characteristics among policy actors, goals, and tools. The findings indicate: (1) Collaboration among core policy actors is close, yet inter-departmental transparency and collaborative inclusivity remain limited for certain organizations. (2) Policy goals show a diversifying trend, with the strategic focus shifting from infectious disease prevention and control to comprehensive public health services. (3) There are significant preferences in the selection of policy tools for balancing rapid emergency response with sustainable long-term health governance. These findings reveal the evolutionary laws of the public health policy system and provide a theoretical basis for optimizing the policy framework and enhancing governance efficacy.

Keywords: public health; policy evolution; social network analysis; topic modeling

1. Introduction

Against the macro-background of a profound restructuring of the global health governance architecture, the "triple crisis" constituted by the pandemic spread of emerging infectious diseases, complex environmental pollution syndromes, and the burden of chronic diseases is reshaping the international public health policy agenda. As a paradigmatic global public health crisis, the spread of COVID-19 has not only exposed the fragility of national public health systems [1] but also highlighted the imperative of perfecting mechanisms for major epidemic prevention and control, as well as strengthening national public health emergency management systems [2]. Constructing a policy framework that balances emergency response efficacy with long-term governance functions under resource constraints remains a pressing challenge in global health governance research [3]. As emphasized by van Reisen et al. [4], the intelligent processing and FAIR (Findable, Accessible, Interoperable, Reusable) management of health data are essential for monitoring crisis responses and optimizing resource allocation.

To address these global issues, China has promulgated a series of public health policies, undergoing a phased transition from an era prioritizing both revolution and hygiene, to one of universal welfare, followed by market-oriented reforms, and finally to the perfection of emergency and health systems [5]. China's public health policy evolution can be broadly categorized into three stages: (1) The infectious disease prevention and control stage (pre-2003), which focused on the construction of epidemiological surveillance systems and the improvement of emergency response mechanisms; (2) The universal health-oriented stage (2003–2015), which realized a transformation toward health promotion strategies through the construction of a series of institutional systems; and (3) The Healthy China construction stage (2015–present), dedicated to building a multi-agent governance pattern involving "government–market–society." This paradigmatic evolution is reflected in legislative milestones such as the Law on Basic Medical and Health Care and the Promotion of Health, and has been fully validated in practices such as the "dynamic zero-COVID" strategy, demonstrating the adaptability and resilience of the institutional system [6].

Despite extensive research on public health policy evolution, existing studies predominantly rely on static content analysis or single-method approaches, which often fail to capture contextual semantic nuances and dynamic multi-agent interactions. Traditional topic modeling techniques, such as Latent Dirichlet Allocation (LDA), operate on bag-of-words assumptions that neglect lexical order and contextual dependencies, thereby obscuring nuanced thematic shifts. Similarly, conventional social network analyses frequently treat policy collaboration as static, overlooking the temporal coupling between institutional restructuring and semantic evolution. The theoretical gap lies in the insufficient integration of advanced natural language processing techniques with network governance theories to elucidate the driving mechanisms behind policy adaptation. Furthermore, current methodologies frequently overlook the validation of topic stability and the structural inertia within collaborative networks, limiting their explanatory power regarding institutional logics. Therefore, this paper constructs a four-dimensional analytical framework comprising "Theme-Actor-Goal-Instrument," integrating the BERTopic model and Social Network Analysis (SNA). Specifically, the BERTopic model is employed to systematically examine the diachronic evolution patterns of policy hotspots within 137 central-level policy documents. Subsequently, Social Network Analysis is utilized to quantify the degree centrality and betweenness centrality of policy actors, thereby deconstructing the evolutionary characteristics of the network structure. Finally, by examining the evolving relationships between policy actors and goals, as well as instruments and goals, this study elucidates the dynamic adaptation mechanism between governance tools and strategic orientation.

We aim to theoretically elucidate the evolutionary mechanisms of public health policy, expand the explanatory boundaries of policy process theory, and provide decision-making references for developing nations seeking to build resilient public health systems.

2. Literature Review

2.1. Institutional Evolution and Governance Logics of Public Health Policy

The development and evolution of China's public health policy have consistently constituted a focal domain within public policy research. Empirical evidence indicates that changes in value consensus serve as the endogenous driving force for policy change, while governmental preferences directly constrain the selection of policy tools [7,8]. Building upon this premise, historical analyses demonstrate that the governing philosophy promoted a structural shift in public health policy from an economic orientation toward a value-driven, health-equity paradigm [9]. In crisis contexts, quantitative modeling has further validated the operational necessity of targeted containment strategies, providing methodological foundations for scientific decision-making during public health emergen-

cies [10]. Recent network topology analyses have revealed the dynamic coordination patterns among multi-source governance subjects during emergency response [11].

Internationally, research on policy response mechanisms has deepened the understanding of organizational and cognitive dimensions in complex governance environments. Reviews of "wicked problems" have highlighted how bureaucratic structures and cognitive biases shape policy formulation and execution [12]. Following the onset of the global pandemic, econometric assessments confirmed that government intervention policies played a decisive role in modulating infection growth rates [13], while cross-national databases emphasized that dynamic policy adjustment remains critical for epidemic control [14]. Despite these contributions, a persistent limitation remains: existing literature predominantly focuses on either macro-level policy shifts or isolated crisis responses, failing to systematically integrate longitudinal thematic evolution with multi-agent collaborative networks.

2.2. Methodological Advances in Policy Text Analysis

Policy texts serve as crucial artifacts for comprehending public policy transitions. Contemporary scholarship has established three primary methodological pathways—structured content analysis, policy bibliometrics, and computational text mining—to examine content structure and evolutionary trajectories [15]. Content analysis emphasizes the systematic coding of keywords, policy actors, and implementation logic. Recent applications have successfully mapped evolutionary pathways in government data-sharing policies [16] and uncovered combinatory mechanisms among policy instruments in crisis scenarios [17]. Bibliometric approaches utilize quantitative mapping to reveal structural characteristics and actor relationships [18], with recent studies evaluating instrument matching deficiencies and inter-departmental synergies in talent evaluation frameworks [19], as well as exploring theoretical–practical interactions in public cultural service policies [20].

With advancements in Natural Language Processing (NLP), computational text mining has gained widespread application. While LDA models have effectively identified phasic characteristics in digital inclusion policies for the elderly [21], PPP policy volatility [22], and regional collaborative policy spillovers [23], their reliance on static word distributions limits contextual interpretability. Conversely, embedding-based models demonstrate superior capacity in capturing latent thematic shifts [24,25]. Recent validations in health informatics confirm that contextual topic modeling outperforms traditional methods in tracing research evolution [26,27], while cluster analysis of large-scale text data effectively identifies key health risk factors and thematic characteristics [28]. Furthermore, Social Network Analysis has emerged as a robust instrument for parsing collaborative structures and topic association networks [29,30]. Scholars have successfully applied co-word and inter-governmental networks to map standardization policy evolution [31], carbon emission reduction collaboration [32], and green shipping governance decentralization [33]. Nevertheless, the integration of contextual topic modeling with dynamic network analysis in public health policy remains underexplored, particularly regarding the validation of semantic stability and the structural mechanisms driving collaborative governance [34].

2.3. Synthesized Research Gaps

Current research targeting public health policy rarely combines advanced contextual topic modeling with longitudinal Social Network Analysis. This methodological silo results in an incomplete understanding of how semantic evolution intersects with network restructuring, leaving the institutional logics and driving mechanisms of policy adaptation inadequately explained. This paper addresses this gap by employing a dual-method approach to systematically examine thematic evolution, instrumental configuration, and actor

interaction mechanisms, thereby contributing to the scientific formulation and synergistic governance of public health policy.

3. Research Design

3.1. Data Sources

We collected 137 public health policy documents issued at the central government level in China from 2003 to 2022. These policy documents were selected from Beida Law Database and the Chinese Government website. Keywords such as "public health," "medical big data," "vaccination," and "health education" were used to retrieve relevant departmental and working documents. Invalid documents, including routine approvals, administrative notices, and bulletins, were excluded, yielding a final sample of 137 policy documents. The evolution of China's public health policies from 2003 to 2022 is divided into two phases: S1 (National Health Promotion Phase, 2003–2015) and S2 (Healthy China Initiative Phase, 2015–2022).

As shown in Figure 1A visually depicts the annual number of policies issued in China from 2003 to 2022, with 2020 seeing the highest number of policy documents released. Figure 1B,C show that during Phase S1, a cumulative total of 85 policy documents were issued, accounting for 62% of all documents released. while Phase S2 there were 52 policy documents issued cumulatively, accounting for 38% of the total.

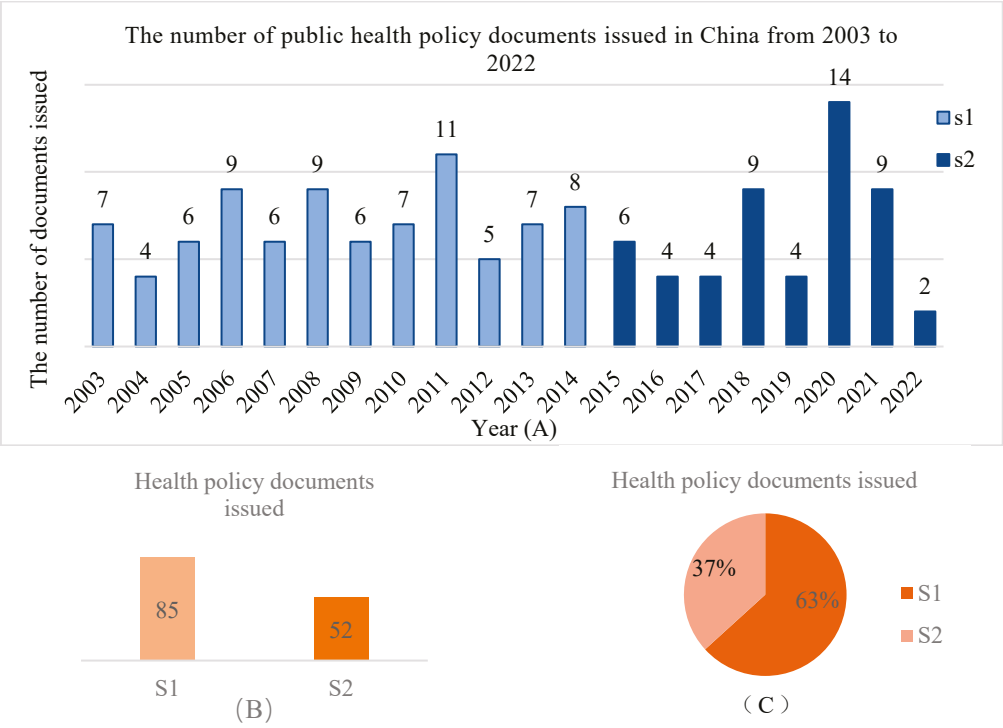


Figure 1. Quantitative distribution of China's public health policy documents (2003–2022). (**A**) The annual volume of policy releases from 2003 to 2022; (**B**) Count of Health policy documents issued during S1 and S2; (**C**) Proportion of health policy documents issued during S1 and S2.

3.2. Research Methods

We employ the BERTopic model to preprocess, embed, and extract topics from policy texts [35]. Based on this, a co-occurrence relationship matrix is constructed for three key policy elements: policy actors, policy objectives, and policy instruments. Using UCINET software for social network analysis [36], then focuses on measuring indicators such as degree centrality, betweenness centrality, and structural holes within the network to

depict the evolutionary structure of the policy network. Furthermore, a visual policy coordination network map was generated to analyze collaborative relationships among policy-making entities, identify core nodes and key pathways within the policy network, thereby revealing structural coordination and power concentration trends during public health policy evolution.

3.2.1. The BERTopic Model

The BERTopic model is employed to analyze the deep thematic structure of policy texts. Its selection is grounded in its capacity to overcome the bag-of-words limitations of traditional models by leveraging contextual embeddings [37]. The procedure involves three steps:

1. Document embeddings are generated using a pre-trained Transformer model, representing each policy as a high-dimensional semantic vector.
2. Dimensionality reduction is performed via UMAP (Uniform Manifold Approximation and Projection), followed by density-based clustering using HDBSCAN to identify cohesive, low-overlap topic clusters.
3. Topic representation is refined using class-based TF-IDF (c-TF-IDF), and is calculated as:

$$W_{t,c} = tf_{t,c} \cdot \log\left(1 + \frac{A}{tf_t}\right) \tag{1}$$

where $tf_{t,c}$ denotes word frequency within cluster c, tf_t represents global frequency, and A is the average document length per cluster. To ensure model reliability, we report a C_v coherence score of 0.62 and a topic diversity index of 0.82, confirming robust semantic interpretability and minimal cluster redundancy [38].

3.2.2. Social Network Analysis (SNA)

Drawing upon network governance theory [39,40], three multi-dimensional matrices are constructed:

(1) Actor Co-occurrence Network (1-mode): nodes represent policy actors, and edge weights quantify joint issuance frequency;
(2) Actor-Goal Association Network (2-mode): maps policy actors to strategic objectives;
(3) Goal-Instrument Mapping Network (2-mode): links governance goals to implementation tools.

UCINET software calculates degree centrality, betweenness centrality, effective size, and constraint. Visualization reveals power concentration trends, collaborative pathways, and structural evolution across phases.

4. Results and Analysis

4.1. Thematic Evolution of Public Health Policy

(1) The Universal Health-Oriented Stage

Thematic modeling was performed on policy documents from the Universal Health-oriented stage (stage S1), identifying a total of 26 topics. The resulting map is presented in Figure 2, where each circle represents a distinct topic. The size of the circle corresponds to the topic's frequency across all documents. As for the distance between circles, it indicates their semantic similarity within the coordinate system. As it can be observed, Topic 0 (Disease Epidemics), Topic 1 (Food Poisoning), and Topic 2 (Mental Health) appear with the highest frequency.

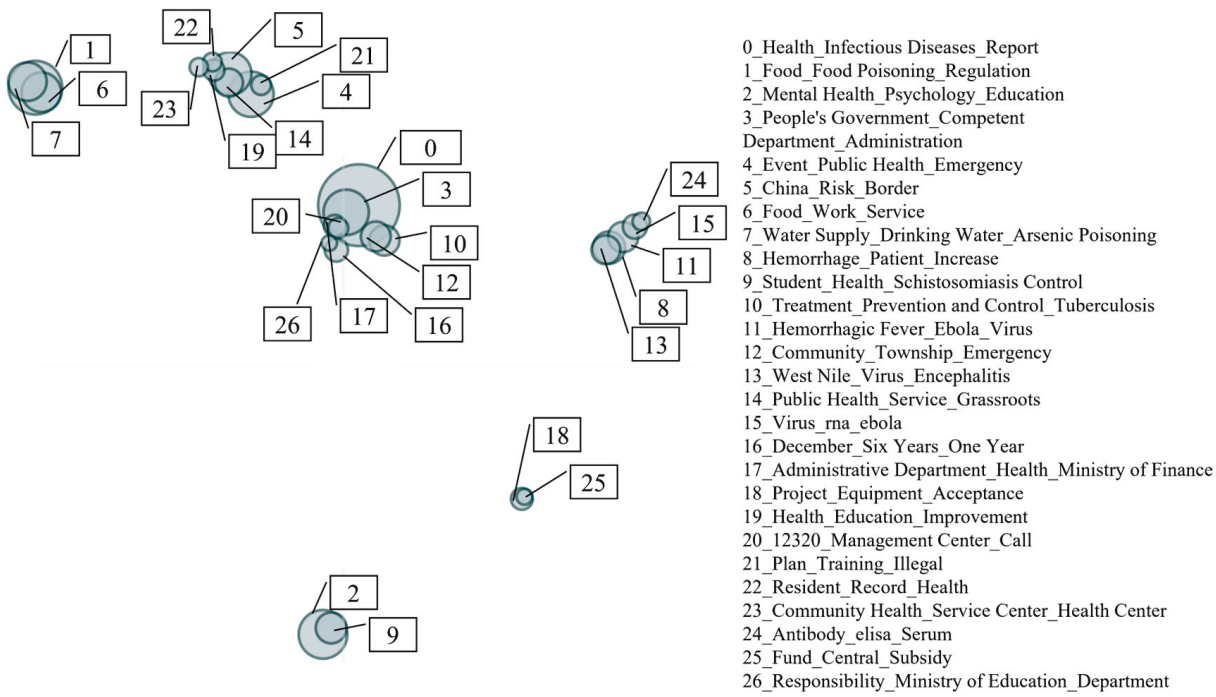


Figure 2. Visualization of the Policy Topic Model for Stage S1.

To comprehend the latent hierarchical structure of these topics, the relationships between keywords were visualized (Figure 3). It intuitively displays associations at various levels. For instance, a direct and close link exists between Topic 1 (Food Poisoning) and Topic 7 (Drinking Water Safety). Similarly, Topic 19 (Health Education) is directly associated with Topic 22 (Health Records), but indirectly associated with Topic 7 (Drinking Water Safety). To illustrate document-topic distributions at a finer granularity, Figure 4 presents a color-coded mapping where each cluster denotes documents assigned to the same topic.

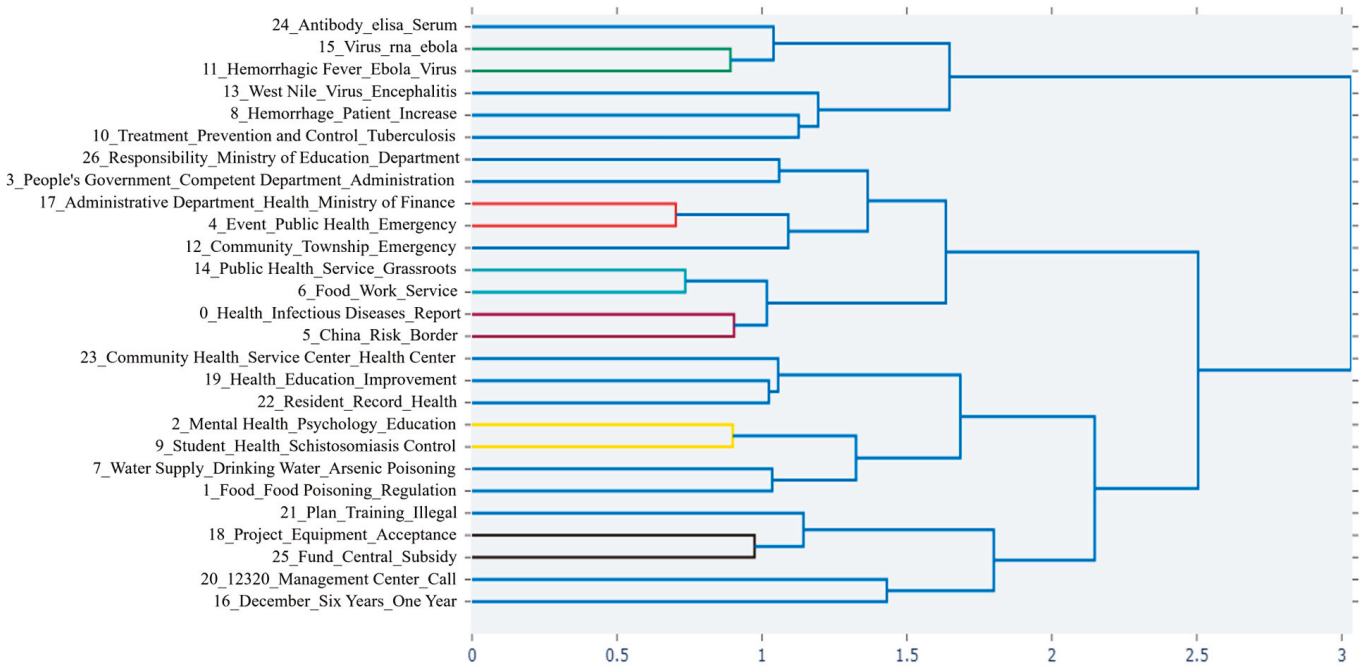


Figure 3. Latent Hierarchical Structure of Policy Topics in Stage S1. Different colors represent distinct clusters of topics that have been grouped based on their similarity.

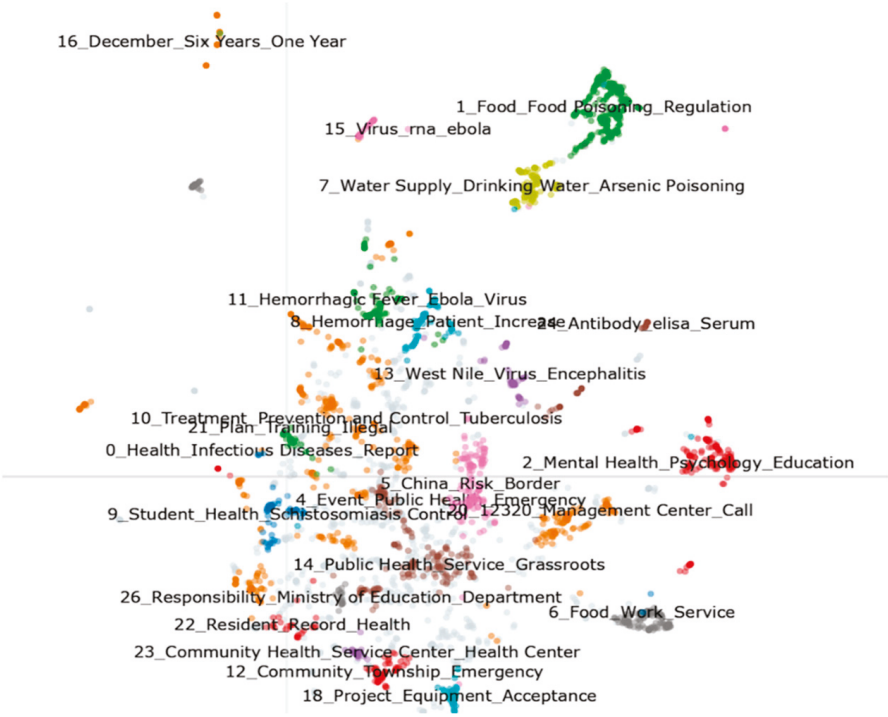


Figure 4. Distribution Map of Policy Text Topics in Stage S1.

(2) The Healthy China Construction Stage

By conducting topic modeling on policy documents from the Healthy China Construction stage (stage S2), a total of 16 topics were identified. The distribution map of policy text is shown in Figure 5, and the relationships between keywords are visualized in Figure 6. Results indicate that Topic 0 (Medical and Health Services) and Topic 1 (Food Poisoning) are the most frequent topics. Furthermore, Figure 6 reveals that Topic 4 (Mental Health) and Topic 7 (Health Education) are directly related, while sharing an indirect connection with Topic 9 (Novel Coronavirus).

Figure 5. Distribution Map of Policy Text Topics in Stage S2.

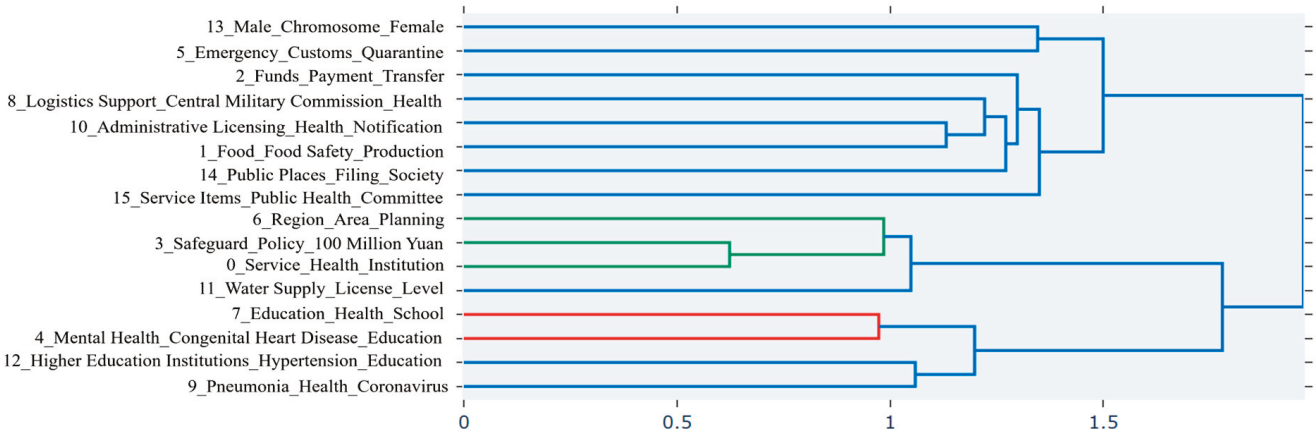


Figure 6. Latent Hierarchical Structure of Policy Topics in Stage S2. Different colors represent distinct clusters of topics that have been grouped based on their similarity, numbers are the topic index.

(3) Comparative Analysis of Stages S1 and S2

The thematic evolution reflects three structural patterns. First, major public health events (SARS in 2003, COVID-19 in 2019) catalyzed abrupt thematic realignment. Second, persistent themes (food safety, drinking water quality, mental health, port quarantine) demonstrate institutional path dependency and sustained governmental prioritization. Third, a strategic shift occurred: Phase S1 emphasized infectious disease control, whereas Phase S2 prioritized comprehensive medical and health services, reflecting a transition from crisis-driven response to proactive health governance. These findings align with prior research on policy lock-in [19], yet extend the literature by demonstrating how digital and service-oriented themes disrupt traditional administrative silos.

4.2. Evolution of Public Health Policy Networks

The evolution of public health policy constitutes a complex network structure. It involves multiple policy-issuing bodies and implementation agencies. Within this structure, policy actors tailor adaptive policy goals and deploy targeted policy instruments across different stages and environmental contexts. Together, these three elements—policy actors, policy goals, and policy instruments—form the dynamic evolutionary network of China's public health policy. This section applies Social Network Analysis (SNA) to this dynamic network to elucidate the structural relationships between policy networks in different phases, identify critical nodes, and measure their functional roles.

(1) Evolution of Policy Actors

As illustrated in Figure 7, during Phase S1, the Ministry of Health and the Ministry of Education accounted for the highest volume of independent policy documents. This underscores their pivotal status and robust functional roles in the institutional construction of China's public health policy during this stage. In Phase S2, however, the State Council's prominence increased significantly. This shift was primarily driven by the issuance of COVID-19 prevention and control policies. The State Council's direct issuance of documents signaled a "speed-is-critical" attitude toward treatment during major public health events and reflected the state's prioritization of life and health in response to the urgency of the crisis.

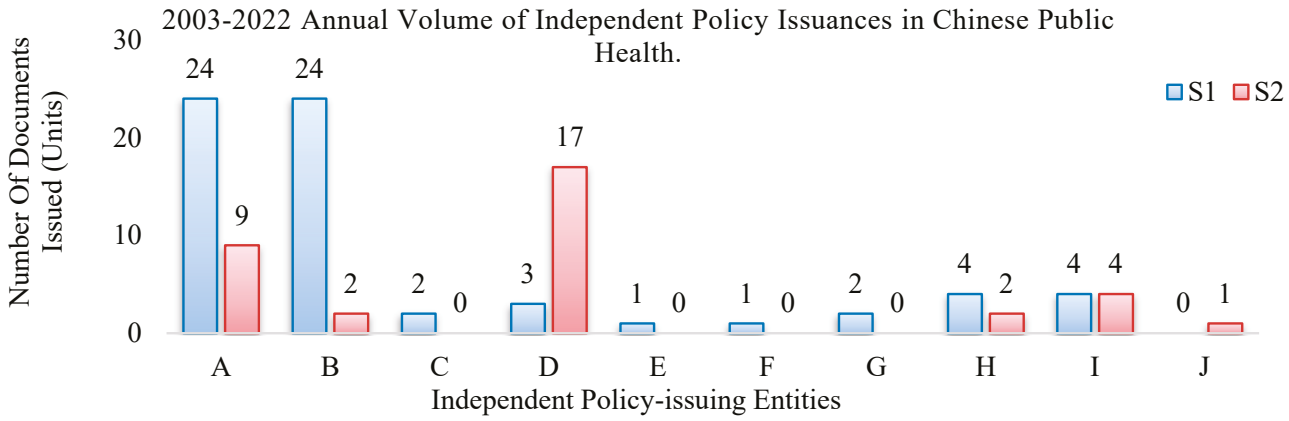


Figure 7. Bar chart of independent issuance volume by policy actors in China's public health policy (2003-2022). (A) Ministry of Health/National Health Commission (MOH/NHC), (B) Ministry of Education (MOE), (C) General Administration of Quality Supervision, Inspection and Quarantine (AQSIQ), (D) State Council, (E) National Headquarters for the Prevention and Control of SARS, (F) National Patriotic Health Campaign Committee (NPHCC), (G) National Administration of Traditional Chinese Medicine (NATCM), (H) State Food and Drug Administration (SFDA), (I) National Health and Family Planning Commission (NHFPC), and (J) State Administration for Market Regulation (SAMR).

Figure 8 demonstrates a shift in authority for independent policy issuance from the Ministry of Health and Ministry of Education in S1 to the State Council in S2. This indicates an escalating level of state emphasis on public health undertakings. Quantitatively, the proportion of policies issued jointly increased from 24% in S1 to approximately 33% of the total in S2, suggesting that S2 placed a greater focus on inter-departmental collaboration.

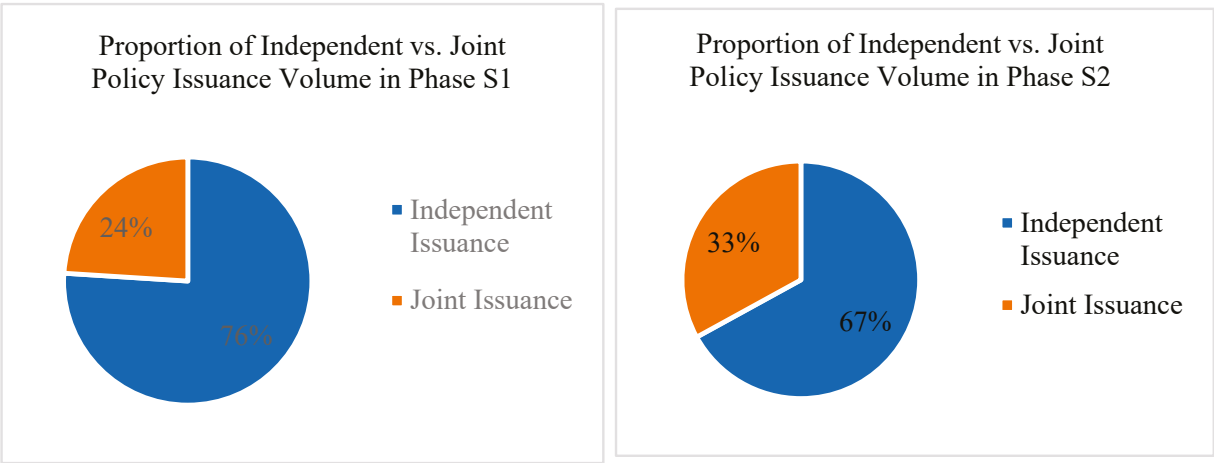


Figure 8. Proportion of independent vs. joint issuance volume in China's public health policy (2003–2022).

(2) Analysis of Collaborative Networks

Analysis of Collaborative Networks: In the policy formulation process, the functional limitations of single actors and the overlapping responsibilities among multiple actors can lead to implementation efficiency reducing or organizational redundancy. To optimize cross-departmental efficacy, national policy-making institutions frequently adopt a joint issuance mode, integrating multi-agent resources to achieve governance goals. We used UCINET software to construct a policy actor synergy matrix, analyzing actor interactions through the dual dimensions of Centrality (Degree Centrality, Betweenness Centrality) and

Structural Holes (Effective Size, Constraint). The dual-centrality topology map in Figure 9 clearly reveals a paradigm shift in the network structure across the two phases. In Global Network Level, the Ministry of Finance consistently occupies the top position in dual-centrality indicators, leading multi-departmental coordination through financial resource allocation. This highlights the strategic leverage of financial assurance mechanisms in policy implementation. Notably, the National Administration of Traditional Chinese Medicine and the National Health Commission (NHC) witnessed a surge in degree centrality during Phase S2. That reflected a significant enhancement in their policy participation frequency and direct association capabilities. Meanwhile, the former Ministry of Health and the National Health and Family Planning Commission (NHFPC) maintained high betweenness centrality due to their administrative hierarchy. It indicated that traditional administrative departments retain the power of information control and resource deployment through institutional authority.

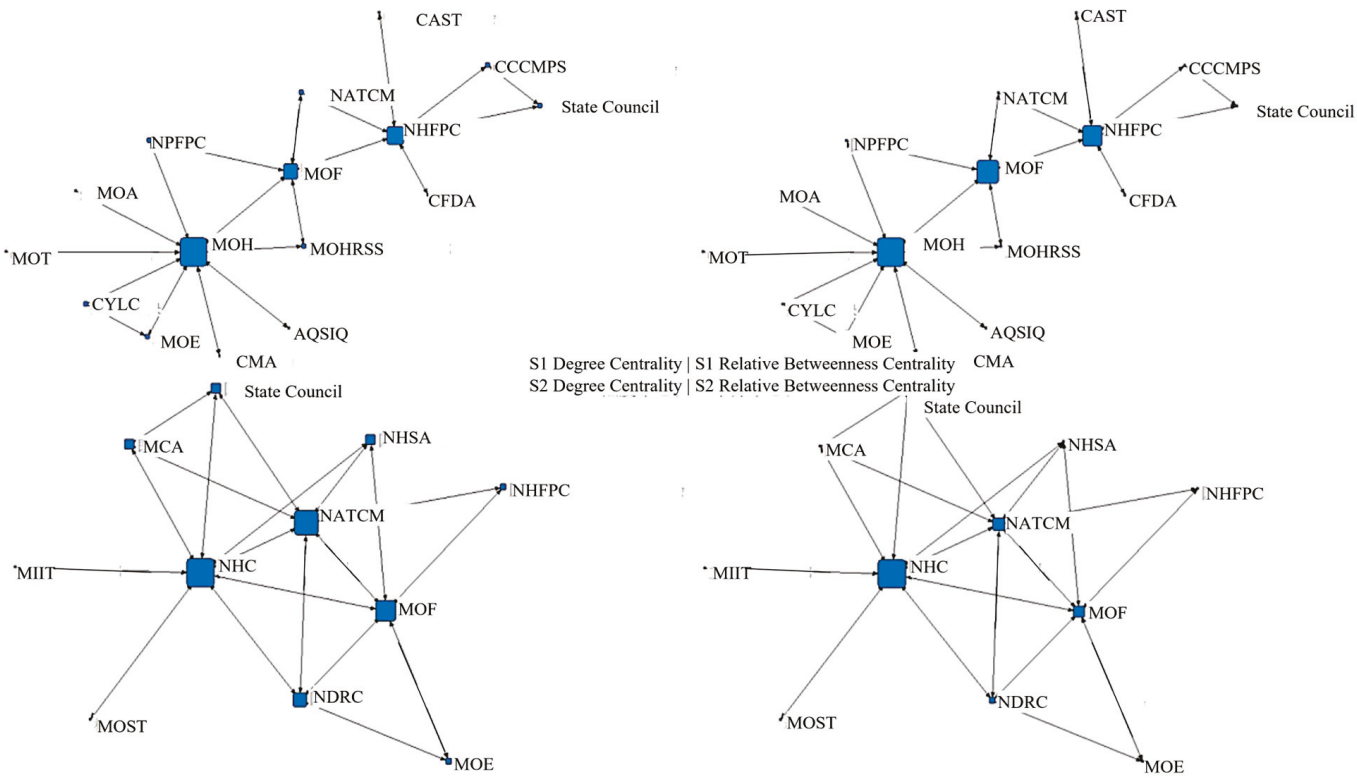


Figure 9. Network topology of joint policy issuance actors in China's public health policy.

Phased structural observations for phase S1 (Monocentric Radiation) show that the collaborative network presented a typical "single-core radiation" structure. The Ministry of Health, with a degree centrality share of 15%, served as the absolute core node. This node implemented vertical control through a triangular support architecture formed with the Ministry of Finance and the NHFPC. 76% of independent issuance rate during this phase corroborates the characteristic of a policy process dominated by a strong centralization mode. As for phase S2 (Polycentric Nesting), the network structure evolved toward a "multi-center nested" morphology. Nodes increased from 12 to 19, and network density increased by 37%, indicating a deep expansion of collaborative governance. The National Administration of Traditional Chinese Medicine, with betweenness centrality of 4.333, emerged as a critical bridge node, propelling the transformation of health governance from single-disease prevention toward an integrated "Prevention–Treatment–Rehabilitation" Big Health paradigm.

Analysis of Structural Holes: Structural hole indicators of policy synergy network can reveal the dynamics of information control among key actors, as shown in Table 1, detailed data are listed Tables A1 and A2. In phase S1, the Ministry of Health and the Ministry of Finance dominated with effective sizes of 4.121 and 3.75, respectively. Their constraint indicators (0.678 and 0.503) constituted a dual-information hub structure which enhanced execution efficiency. However, 81% of nodes exceeded theoretical constraint threshold (0.7). This indicated a severe insufficiency in network openness and significant institutional barriers between departments. In phase S2, the network exhibits a structural transformation. The National Administration of Traditional Chinese Medicine became primary occupant of structural holes with effective size of 4.095, while its constraint decreased by 30% compared to S1. Meanwhile, Cross-departmental collaborative formed with the NHC, which drove the transformation of the governance paradigm. Conversely, though the State Council had a 24% increase in effective size, its constraint decreased by 38%. From overall evolutionary trends perspective, network density increased from 0.134 to 0.274, indicating a significant expansion in range of cooperation among policy actors. But, mean constraint value decreased by only 8.6%, exposing a structural inertia characterized by "broad connection but low efficiency." Core nodes, like Ministry of Finance and NHC, maintain resource control advantages through high constraint. However, peripheral actors, like Ministry of Education, are limited by a constraint value of 1.125, which struggling to breach structural constraints to participate in deep collaboration. This persistent "Core–Periphery" differentiation suggests that dissolution of information barriers requires in-depth institutional innovation.

Table 1. Effective Size and Constraint of Joint Issuance Actors in China's Public Health Policy (Partial Data).

No.	Policy Entity	Effective Size			Degree of Restriction		
		S1-2	S1	S2	S1-2	S1	S2
1	Ministry of Finance (MOF)	3	3.75	3.556	0.535	0.503	0.706
2	National Administration of Traditional Chinese Medicine (NATCM)	1.667	1.528	4.095	0.84	0.918	0.64
3	National Health Commission (NHC)	2.333	-	3.256	0.611	-	0.777
4	Ministry of Health (MOH)	3	4.121	-	0.333	0.678	-
5	National Health and Family Planning Commission (NHFPC)	1	4.335	1	1.125	0.559	1.531
6	Ministry of Education (MOE)	1	1.152	1	1	0.888	1.125
7	State Council	-	1.267	1.571	-	0.837	1.156

(3) Evolution of Policy Actor–Policy Goal Relationships

A two-mode network model of "Policy Actors–Policy Goals" was constructed based on multi-agent governance framework of Chinese public health policy. The network systematically elucidates the evolutionary of governance system from 2003 to 2022. Results are illustrated in Figure 10.

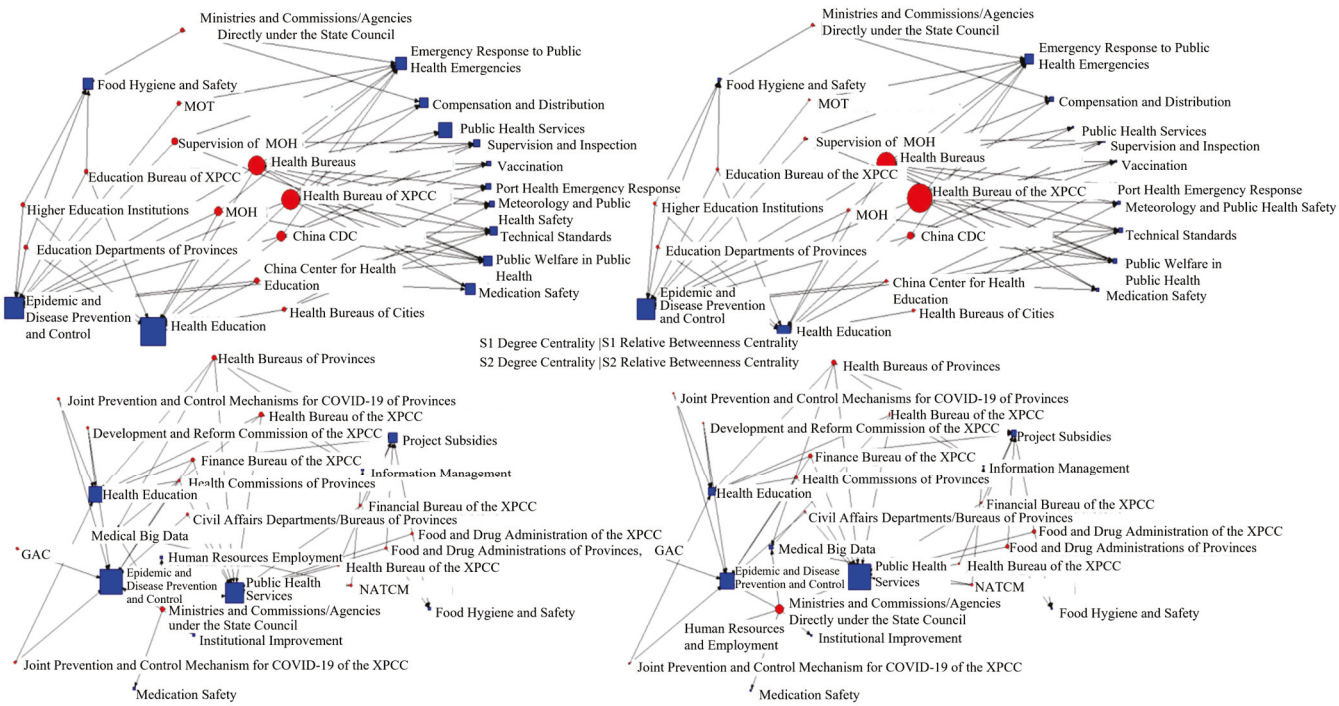


Figure 10. Network Topology of "Policy Actors–Policy Goals" in Chinese Public Health Policy. The visualization uses blue squares to indicate policy goals and red dots for policy actors. The size of each node is proportional to its centrality, highlighting the relative importance within the network.

Regarding the actor dimension, the number of policy nodes expanded from 15 in Phase S1 to 23 in Phase S2, indicating continuous enhancement in governance system specialization, detailed data are listed Tables A3 and A4. In Phase S1, the network exhibited a "dual-core drive" pattern: provincial education departments advanced health promotion through multi-departmental collaboration, while health departments dominated agenda setting via policy resource control. Notably, although the Chinese Center for Disease Control and Prevention exhibited low collaborative frequency (degree centrality: 0.035), it retained critical discursive power in professional decision-making (betweenness centrality: 2.173). In Phase S2, the integration of the Joint Prevention and Control Mechanism for COVID-19 significantly reconfigured the network. Its average degree centrality (1.87) surpassed that of conventional governance actors, validating the coupling between bureaucratic administration and campaign-style governance during major crises.

From the goal dimension perspective, the Phase S1 network centered on epidemic prevention and control, radiating to other domains through institutional emergency response channels. In Phase S2, a structural leap occurred: "Public Health Services" ascended to the core position, and the network expanded to encompass the 15 specialized Healthy China Actions, covering emerging areas such as elderly care and medical-prevention synergy. This reflects a paradigm shift from "passive emergency response" to "proactive health governance." Concurrently, emerging goals like medical big data and human resource development achieved strategic embedding, reflecting post-pandemic digital transformation and a renewed focus on health equity. However, 78% of nodes exhibited zero betweenness centrality, exposing structural contradictions characterized by core–periphery differentiation and low synergistic efficiency. This underscores the urgent need for institutional innovation to dismantle resource monopolies and enhance participatory governance.

(4) Evolution of Policy Goal–Policy Instrument Relationships

Based on policy instrument theory, a "Goal–Instrument" two-mode network model was constructed. A relationship matrix was established by selecting 50 policy instruments

and 24 policy goals from the period 2003–2022. The matrix systematically parsed the dynamic adaptation mechanism of China's public health policy system, detailed data are listed Tables A5 and A6. Results are shown in Figure 11.

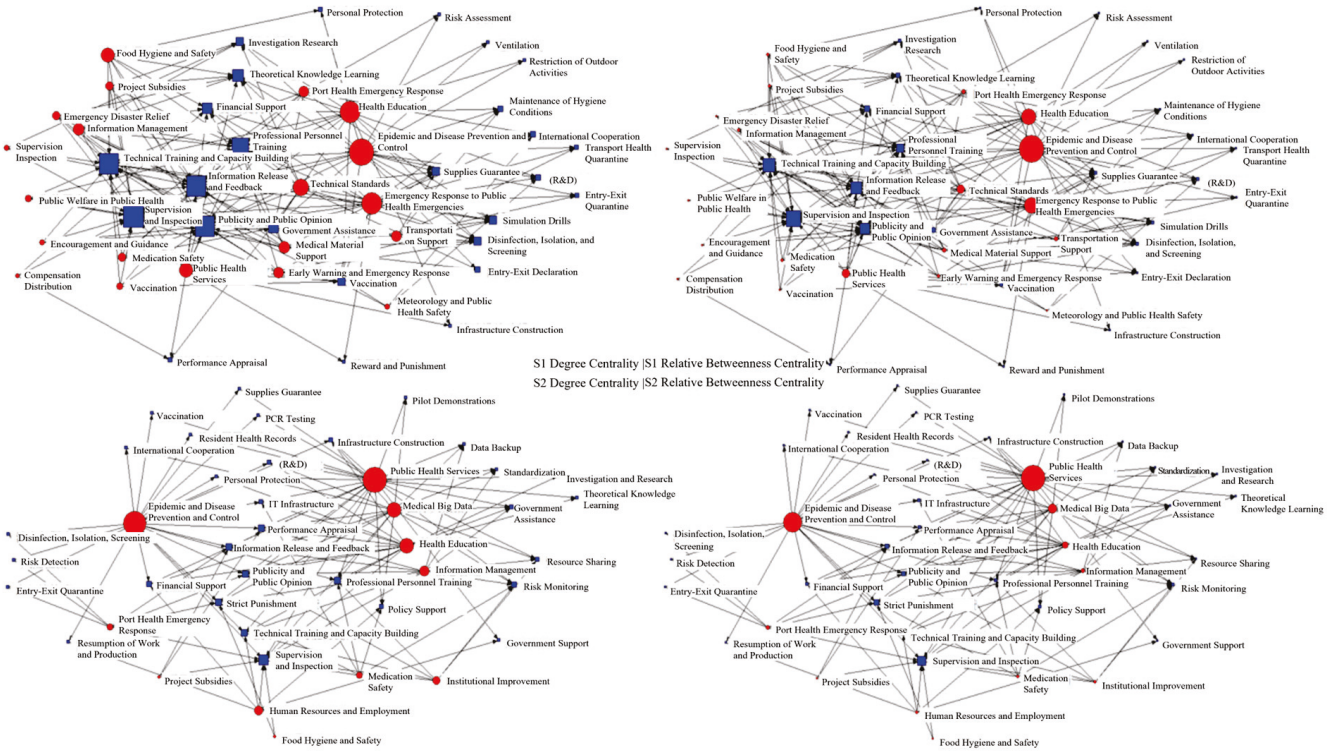


Figure 11. Network Topology Map of Policy Goals and Policy Tools in China's Public Health Policy. The visualization uses blue squares to indicate policy tools and red dots for policy goals. The size of each node is proportional to its centrality, highlighting the relative importance within the network.

In phase S1, the policy instrument network featured a core tool cluster include Supervision and Inspection, Propaganda and Mobilization, and Information Release and Feedback (mean degree centrality 0.105, relative betweenness centrality > 2.5%). In phase S2, instrument innovation exhibited a "Three-Track Parallel" characteristic. First is Integration and Proliferation of Digital Tools. In detail, the degree centrality of Informatization Construction rose to 0.053. And the embedding of resource sharing and data backup increased significantly, illustrate a transition in policy execution toward intelligent decision-making. Second is Specialization of Emergency Tools. Specifically, eight COVID-specific tools, such as Nucleic Acid Testing and Mask Mandates etc., formed independent functional modules, highlighting response capabilities for major public health events. Last is Supplementation of Livelihood Assurance Tools. It includes thirteen economic tools, like Work Resumption and Job Stabilization Subsidies etc., which filled traditional gaps. Their mean degree centrality of 0.04 reflects high sensitivity of policy system to socioeconomic demands.

The quantitative transformation of the policy actor network cannot be interpreted merely as a statistical expansion of collaborative ties; rather, it reflects a profound recalibration of institutional governance logic [41]. The transition from a monocentric to a polycentric architecture demonstrates the state's strategic response to the functional limitations of departmental silos under complex health challenges. The persistence of high betweenness centrality and structural constraint among core entities indicates that policy coordination remains fundamentally anchored in bureaucratic resource allocation and institutional authority. Consequently, the evolution of the public health policy network adheres to a governance logic wherein collaborative efficiency is continuously negotiated against

path-dependent administrative boundaries and information asymmetries. This structural inertia necessitates institutional innovations that transcend temporary crisis mobilization, transforming ad hoc coordination into standardized, rule-based inter-departmental governance mechanisms.

4.3. Case Study

The central-level policy framework provides a unified institutional foundation for the whole country, yet differences in economic development levels, resource endowments, and governance capacities across regions lead to significant heterogeneity in policy implementation. To reveal this phenomenon, we selected Guangdong, a developed coastal province, and Sichuan, a less developed inland province, as comparative cases. Focusing on the COVID-19 pandemic period (2020–2022), which falls within Phase S2 of this study, we compare the local implementation of policies in the two provinces across three dimensions: policy goals, policy instruments, and actor collaboration patterns, as shown in Table 2. Data for the two provinces are derived from provincial-level COVID-19 policy documents issued between January 2020 and December 2022, primarily collected from the official websites of the two provinces' health commissions, provincial government portals, and authoritative media reports.

Table 2. Comparative Case Analysis of Guangdong and Sichuan Provinces.

Dimension	Guangdong Province	Sichuan Province
Policy Instruments	Highly dominated by informatization tools	Balancing informatization with grid-based management
Actor Collaboration Patterns	Multi-center parallel collaboration type	Multi-center radial collaboration type
Policy Goals	Coordinated health and economic development	Primary health protection

In terms of policy instruments, the central-level analysis reveals that Phase S2 is characterized by the rise of informatization construction tools and the diversification of instrument types, exhibiting a "three-track parallel" feature. Both provinces demonstrate the application of informatization tools. However, leveraging its higher level of informatization infrastructure, Guangdong has achieved a technology-driven rapid transition dominated by digital tools such as the Yuekang Code, big-data contact tracing, and grid-based digital management. In contrast, constrained by limited resources, Sichuan relies on organizational mobilization, such as grid staff conducting door-to-door outreach and "knock-on-door actions", as the main support for the implementation of digital tools, forming a pathway that balances informatization with grid-based management. Both align with the overall trend of instrument diversification, with the difference originating from regional disparities in informatization levels.

In terms of actor collaboration patterns, the central-level analysis finds that the joint issuance network of policy actors in Phase S2 evolved from the "single-core radiation" structure of Phase S1 toward a "multi-center nested" morphology. Following this trend, both provinces also exhibit multi-actor collaborative features at the local implementation level, yet with distinct patterns. Guangdong presents a "multi-center parallel collaboration type", in which multiple actors participate in parallel within the network through mechanisms such as rural COVID-19 prevention task forces and "three-person groups". on the other hand, Sichuan exhibits a "multi-center radial collaboration type". At the institutional design level, Sichuan explicitly designates the health department as the lead, while embedding the responsibilities of public security, transportation, industry and information technology, commerce, education, and propaganda departments into the emergency response process,

forming a regularized division of labor and collaboration list. Vertically, it strengthens the professional guidance and unified coordination of the provincial, municipal, and county command systems.

In terms of policy goals, the central-level analysis shows that the strategic focus of Phase S2 shifted from "prevention and control of infectious epidemics" in Phase S1 to "medical and health services". Under this macro framework of "medical and health services", the two provinces have developed differentiated sub-goal priorities. As a major export-oriented economy, Guangdong prioritizes "coordinated health and economic development", implementing economic safeguards through measures such as facilitating work resumption and production via the Yuekang Code and introducing enterprise support policies. As a major labor-exporting province, Sichuan prioritizes "primary health protection", focusing on grassroots control through measures such as the "four-level responsibility contracting" system and the "five haves and one grid" approach. These differences represent adaptive adjustments of the macro goal framework based on local endowments, and neither of them deviates from the overall strategic direction of Phase S2.

5. Conclusions and Outlook

5.1. Core Findings and Institutional Logics

This study employs a dual methodological approach combining BERTopic modeling and Social Network Analysis to systematically examine the evolutionary trajectory of China's public health policy from 2003 to 2022. The empirical analysis reveals that China's public health policy evolution is not a linear accumulation of regulatory documents, but a dynamic process governed by the interplay of exogenous shocks, endogenous bureaucratic restructuring, technological advancement, and institutional path dependence. To transcend descriptive summaries, this study constructs an analytical framework for the driving mechanisms of public health policy evolution, elucidating four core governance logics.

Crisis-induced adaptation serves as the primary exogenous driver catalyzing thematic paradigm shifts and institutional punctuated equilibrium. The recurrent prominence of infectious disease control and emergency response in the policy corpus demonstrates how macro-level public health emergencies disrupt institutional inertia, forcing rapid agenda reconfiguration and resource reallocation. The post-2020 emergence of digital health and medical big data topics illustrates a systemic transition from reactive crisis management to proactive, data-driven health governance [42]. This mechanism underscores that policy evolution in transitional contexts is fundamentally characterized by non-linear acceleration, where external shocks catalyze the institutionalization of latent policy preferences and compress the timeline for structural adaptation.

Bureaucratic resource allocation and power restructuring constitute the endogenous mechanism governing collaborative network dynamics and inter-agency synergy. The observed structural transformation from a monocentric radiation model to a polycentric nested architecture reflects the state's strategic recalibration of administrative authority to overcome functional fragmentation. The persistence of high structural constraint and central resource control among core entities, such as the Ministry of Finance and the National Health Commission, reveals a governance logic wherein financial assurance and professional expertise function as critical leverage points for cross-sectoral coordination. Consequently, policy network evolution remains inherently bounded by bureaucratic politics, wherein collaborative expansion must continuously negotiate institutional path dependencies, authority distribution, and information control asymmetries.

Technological empowerment and instrument modernization drive the operational transformation of policy implementation and governance capability enhancement. The structural shift from traditional administrative commands toward intelligent, data-centric

instruments highlights the state's systematic pursuit of technical governance capacity. The integration of digital monitoring platforms, inter-departmental resource-sharing protocols, and targeted economic subsidies demonstrates a pragmatic alignment of policy tools with complex, multi-dimensional governance objectives. However, the enduring structural imbalance in instrument configuration indicates that technological adoption remains functionally compartmentalized; without systemic integration and normative standardization, digital instruments risk operating as supplementary measures rather than foundational governance infrastructure, thereby constraining long-term institutional resilience.

Institutional path dependence and bounded policy learning shape the longitudinal continuity and adaptive refinement of the public health governance system. The sustained policy emphasis on foundational public health services, health education, and food safety across both evolutionary phases evidences a strong path-dependent trajectory rooted in developmental state priorities and foundational welfare commitments. Simultaneously, the iterative adjustment of actor-goal-instrument alignments reflects a process of incremental policy learning, wherein institutional actors systematically refine governance strategies based on empirical performance feedback and evolving societal health demands. This dual dynamic of structural continuity and strategic adaptation ensures that policy evolution maintains institutional stability while accommodating emergent public health challenges.

5.2. Practical Implications

The evolutionary trajectory of China's public health policy adheres to three fundamental governance laws: (1) the law of crisis-driven institutional punctuated equilibrium, wherein exogenous public health shocks accelerate structural transformation and compress policy learning cycles; (2) the law of resource-constrained bureaucratic synergy, which dictates that network expansion and collaborative density remain bounded by institutional capacity, fiscal leverage, and authority distribution; and (3) the law of instrument-goal dynamic congruence, emphasizing that long-term policy efficacy depends on the continuous, systemic alignment of governance tools with strategic objectives rather than scenario-specific improvisation. To optimize future public health governance, policymakers must institutionalize crisis-derived coordination mechanisms, dismantle structural barriers to cross-departmental data and resource sharing, and cultivate a balanced instrument portfolio that integrates short-term responsiveness with long-term systemic resilience.

5.3. Limitations and Future Research Directions

We acknowledge several limitations. First, the temporal scope (2003–2022) excludes post-pandemic policy normalization and long-term structural adjustments emerging after 2022. Second, the sample is restricted to central-level documents, omitting provincial heterogeneity and implementation variations. Future research should incorporate multi-level comparative studies to examine regional adaptation mechanisms. Third, while BERTopic and SNA provide robust semantic and structural insights, they do not capture implementation outcomes or grassroots feedback. Subsequent studies should integrate mixed-method designs, combining text analytics with field surveys and policy execution tracking. Finally, dynamic causal modeling and machine learning-driven policy simulation should be employed to forecast institutional evolution under varying crisis scenarios, thereby enhancing the predictive capacity and systemic resilience of public health governance frameworks.

Author Contributions: Conceptualization L.S. and D.G.; methodology, L.S. and K.Z.; data curation and formal analysis, L.S. and K.Z.; investigation and visualization, K.Z. and W.L.; writing—original draft preparation, K.Z. and W.L.; writing—review and editing, K.Z., X.Y. and W.L.; supervision D.G. and X.Y.; project administration D.G.; funding acquisition, D.G. and X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant Nos. 72131006, 72271082, 72071063, 72571086, 72501088) and the Natural Science Foundation of Anhui Province (Grant No. 2408085J041).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable. We would like to clarify that our study is based on publicly available online consultation data obtained from a widely used Chinese online healthcare platform, with all data processing conducted in compliance with the platform's terms of service and relevant ethical guidelines. Therefore, informed consent was not required.

Data Availability Statement: The original data presented in paper are openly available at https://doi.org/10.6084/m9.figshare.31671634, accessed on 12 March 2026.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Degree Centrality and Betweenness Centrality of Joint Policy-Issuing Entities in China's Public Health Policy.

No.	Policy Entity	Degree Centrality			Betweenness Centrality		
		S1-2	S1	S2	S1-2	S1	S2
1	Ministry of Finance (MOF)	30	9	21	177.367	109	5.833
2	National Administration of Traditional Chinese Medicine (NATCM)	25	2	23	34.55	-	7.833
3	National Health Commission (NHC)	18	-	18	22.55	-	4.333
4	Ministry of Health (MOH) [Historical]	15	15	-	147.75	137	-
5	National Health and Family Planning Commission (NHFPC) [Historical]	14	8	6	99.233	98	-
6	Ministry of Education (MOE)	6	4	2	18.25	-	-
7	State Council	5	2	3	12	-	-
8	National Development and Reform Commission (NDRC)	5	-	5	6.3	-	2
9	National Healthcare Security Administration (NHSA)	3	-	3	-	-	-
10	Ministry of Civil Affairs (MCA)	3	-	3	-	-	-
11	General Administration of Quality Supervision, Inspection and Quarantine (AQSIQ) [Historical]	2	2	-	-	-	-
12	National Population and Family Planning Commission (NPFPC) [Historical]	2	2	-	-	-	-
13	Central Committee of the Communist Youth League (CYLC)	2	2	-	-	-	-
14	Central Committee for Comprehensive Management of Public Security [Historical]	2	2	-	-	-	-
15	Ministry of Human Resources and Social Security (MOHRSS)	2	2	-	-	-	-
16	China Food and Drug Administration (CFDA) [Historical]	1	1	-	-	-	-
17	China Association for Science and Technology (CAST)	1	1	-	-	-	-
18	Ministry of Agriculture (MOA) [Historical]	1	1	-	-	-	-

Table A2. Effective Size and Constraint of Joint Policy-Issuing Entities in China's Public Health Policy.

No.	Policy Entity	Effective Size			Constraint (or Degree of Constraint)		
		S1-2	S1	S2	S1-2	S1	S2
1	Ministry of Finance (MOF)	3	3.75	3.556	0.535	0.503	0.706
2	National Administration of Traditional Chinese Medicine (NATCM)	1.667	1.528	4.095	0.84	0.918	0.64
3	National Health Commission (NHC)	2.333	-	3.256	0.611	-	0.777
4	Ministry of Health (MOH) [Historical]	3	4.121	-	0.333	0.678	-
5	National Health and Family Planning Commission (NHFPC) [Historical]	1	4.335	1	1.125	0.559	1.531
6	Ministry of Education (MOE)	1	1.152	1	1	0.888	1.125
7	State Council	-	1.267	1.571	-	0.837	1.156
8	National Development and Reform Commission (NDRC)	1	1	2.429	1	1	0.913
9	National Healthcare Security Administration (NHSA)	-	1	1.233	-	1	1.253
10	Ministry of Civil Affairs (MCA)	-	-	1.571	-	-	1.156
11	General Administration of Quality Supervision, Inspection and Quarantine (AQSIQ) [Historical]	1	1	-	1	1.125	-
12	National Population and Family Planning Commission (NPFPC) [Historical]	-	1	-	-	1.62	-
13	Central Committee of the Communist Youth League (CYLC)	-	1	-	-	1.531	-
14	Central Committee for Comprehensive Management of Public Security [Historical]	-	1	-	-	1.125	-
15	Ministry of Human Resources and Social Security (MOHRSS)	-	1	-	-	1.62	-
16	China Food and Drug Administration (CFDA) [Historical]	-	1.15	-	-	0.94	-
17	China Association for Science and Technology (CAST)	-	1	-	-	1	-
18	Ministry of Agriculture (MOA) [Historical]	-	1	-	-	1	-

Table A3. Degree Centrality and Betweenness Centrality of Policy Entities in China's Public Health Policy.

No.	Policy Entity	Degree Centrality			Relative Betweenness Centrality (%)		
		S1-2	S1	S2	S1-2	S1	S2
1	Health Bureaus of Provinces, Autonomous Regions, and Municipalities [Historical]	0.108	0.116	0.065	8.123	20.047	2.161
2	Education Departments of Provinces, Autonomous Regions, and Municipalities	0.089	0.157	0.008	0.752	1.364	-
3	Health Bureau of the XPCC	0.074	0.096	0.042	12.161	26.814	2.161
4	Education Bureau of the XPCC	0.068	0.122	0.008	0.752	1.364	-

Table A3. *Cont.*

No.	Policy Entity	Degree Centrality			Relative Betweenness Centrality (%)		
		S1-2	S1	S2	S1-2	S1	S2
5	Health Commission of the XPCC	0.045	-	0.085	2.085	-	5.263
6	Health Commissions of Provinces, Autonomous Regions, and Municipalities	0.045	-	0.085	2.085	-	5.263
7	Finance Departments/Bureaus of Provinces, Autonomous Regions, and Municipalities	0.047	0.018	0.078	2.42	0.56	5.263
8	Administrations of Traditional Chinese Medicine of Provinces, Autonomous Regions, and Municipalities	0.039	0.007	0.079	0.919	-	1.57
9	Higher Education Institutions Affiliated with Ministries	0.027	0.053	0.003	0.752	1.364	-
10	Ministries and Commissions/Agencies Directly under the State Council	0.026	0.004	0.049	3.992	2.622	1.57
11	People's Governments of Provinces, Autonomous Regions, and Municipalities	0.026	0.004	0.049	2.69	0.624	1.57
12	Financial Bureau of the XPCC	0.022	0.01	0.031	0.335	0.56	-
13	Chinese Center for Disease Control and Prevention (China CDC)	0.028	0.035	0.023	2.693	2.173	1.57
14	Finance Bureau of the XPCC	0.021	-	0.051	2.085	-	5.263
15	Administration of Traditional Chinese Medicine of the XPCC	0.019	-	0.046	0.919	-	1.57
16	Direct-Affiliated Inspection and Quarantine Bureaus	0.011	0.021	-	0.054	0.036	-
17	Population and Family Planning Commission of the XPCC	0.009	0.008	-	0.357	0.627	-
18	Food and Drug Administrations of Provinces, Autonomous Regions, and Municipalities	0.009	0.011	0.01	1.399	0.914	-
19	China CDC (Abbreviated)	0.008	-	0.02	0.919		1.57
20	Ministry of Health (MOH) [Historical]	0.006	0.014	-	1.292	1.617	-
21	Health Bureaus of Cities Specifically Designated in the State Plan	0.006	0.013	-	0.262	0.505	-
22	Population and Family Planning Commissions of Provinces, Autonomous Regions, and Municipalities	0.006	0.006	-	0.357	0.627	-
23	China Center for Health Education	0.006	0.012	-	2.286	4.33	-
24	Food and Drug Administration of the XPCC	0.006	0.005	0.01	0.234	-	-
25	Healthcare Security Administration of the XPCC	0.006	-	0.014	0.919	-	1.57
26	Civil Affairs Departments/Bureaus of Provinces, Autonomous Regions, and Municipalities	0.009	0.003	0.012	0.919	-	1.57
27	Ministry of Education (MOE)	0.005	0.012	-	-	-	-
28	Health News (Press/Newspaper)	0.005	0.011	-	0.45	0.221	-
29	People's Medical Publishing House	0.005	0.011	-	0.45	0.221	-

https://doi.org/10.3390/systems14050497

Table A3. *Cont.*

No.	Policy Entity	Degree Centrality			Relative Betweenness Centrality (%)		
		S1-2	S1	S2	S1-2	S1	S2
30	China Population Communication Center (CPCC)	0.005	0.011	-	0.45	0.221	-
31	China Population News	0.005	0.011	-	0.45	0.221	-
32	China Population Publishing House	0.005	0.011	-	0.45	0.221	-
33	Development and Reform Commission of the XPCC	0.005	-	0.008	0.45		2.483
34	Joint Prevention and Control Mechanisms for COVID-19 of Provinces, Autonomous Regions, and Municipalities	0.005	-	0.008	0.45	-	2.483
35	Joint Prevention and Control Mechanism for COVID-19 of the XPCC	0.005	-	0.008	0.45	-	2.483
36	Health Supervision Center of the Ministry of Health	0.004	0.009	-	0.864	1.066	-
37	Education Commission of the XPCC	0.003	0.006	-	-	-	-
38	Decree of the Ministry of Transport of the PRC [Likely a citation error in original source, usually denotes a document]	0.003	0.006	-	0.215	0.345	-
39	Secondary Vocational Schools	0.003	0.006	-	-	-	-
40	Central Committee of the Communist Youth League (CYLC)	0.003	0.006	-	-	-	-
41	Health Administrative Departments of the State Council	0.003	0.006	-	0.042	-	-
42	Education Departments/Bureaus of Relevant Departments (Units)	0.003	0.008	-	0.193	0.914	-
43	Patriotic Health Campaign Committees (Offices) of Provinces, Regions, Municipalities, and the XPCC	0.003	0.003	-	-	-	-
44	China Center for Health Education	0.003	0.007	-	0.013	0.033	-
45	Education Departments (Bureaus) of Relevant Departments (Units)	0.003	0.006	-	-	-	-

Table A4. Degree Centrality and Betweenness Centrality of Policy Goals in China's Public Health Policy for evolution of policy actor–policy goal.

No.	Policy Goal (Objective)	Degree Centrality			Relative Betweenness Centrality (%)		
		S1-2	S1	S2	S1-2	S1	S2
1	Technical Standards	0.028	0.038	-	1.285	1.285	-
2	Epidemic and Disease Prevention and Control	0.164	0.189	0.14	8.74	8.74	24.444
3	Health Education	0.137	0.158	0.094	3.277	3.277	1.111
4	Public Health Services	0.129	0.076	0.209	1.856	1.856	21.111
5	Information Management	0.048	0.043	0.023	-	-	-

Table A4. *Cont.*

No.	Policy Goal (Objective)	Degree Centrality			Relative Betweenness Centrality (%)		
		S1-2	S1	S2	S1-2	S1	S2
6	Project Subsidies	0.034	0.018	0.05	0.445	0.445	6.667
7	Food Safety and Hygiene	0.03	0.038	0.006	0.226	0.226	-
8	Supervision and Inspection	0.028	0.037	-	-	-	-
9	Emergency Response to Public Health Emergencies	0.02	0.024	-	5.479	5.479	-
10	Emergency Disaster Relief	0.019	0.025	-	-	-	-
11	Public Welfare in Public Health	0.016	0.022	-	2.569	2.569	-
12	Port Health Emergency Response	0.015	0.02	0.001	1.285	1.285	-
13	Medication Safety	0.012	0.013	0.013	1.856	1.856	-
14	Medical Big Data	0.012	-	0.026	1.856	1.856	8.889
15	Compensation and Distribution	0.011	0.013	-	1.856	1.856	-
16	Medical and Material Support	0.01	0.013	-	1.285	1.285	0
17	Meteorology and Public Health Safety	0.009	0.013	-	-	-	-
18	Vaccination	0.004	0.006	-	-	-	-
19	Institutional Improvement	0.002	-	0.013	-	-	-
20	Human Resources and Employment	0.002	-	0.013	-	-	-
21	Early Warning and Emergency Response	0.001	0.001	-	-	-	-
22	Encouragement and Guidance	0.001	0.001	-	-	-	-
23	Transportation Support	0.001	0.001	-	-	-	-

Table A5. Degree Centrality and Betweenness Centrality of Policy Goals in China's Public Health Policy for evolution of policy goal–policy tool.

No.	Policy Goal (Objective)	Degree Centrality			Relative Betweenness Centrality (%)		
		S1-2	S1	S2	S1-2	S1	S2
1	Epidemic and Disease Prevention and Control	0.189	0.191	0.176	0.634	0.673	-
2	Health Education	0.164	0.18	0.121	0.634	0.673	-
3	Public Health Services	0.135	0.07	0.222	0.634	0.673	-
4	Food Hygiene and Safety	0.054	0.068	0.021	0.634	0.673	-
5	Emergency Response to Public Health Emergencies	0.04	0.056	-	0.634	0.673	-
6	Information Management	0.037	0.039	0.032	0.634	0.673	-
7	Technical Standards	0.029	0.041	-	0.634	0.673	-
8	Port Health Emergency Response	0.022	0.023	0.018	0.019	0.029	-
9	Project Subsidies	0.02	0.016	0.028	0.634	0.673	-
10	Medical and Material Support	0.017	0.024	-	0.019	0.029	-
11	Public Welfare in Public Health	0.014	0.02	-	0.019	0.029	-
12	Medication Safety	0.013	0.013	0.014	0.019	0.029	-
13	Emergency Disaster Relief	0.01	0.014	-	0.634	0.673	-

Table A5. *Cont.*

No.	Policy Goal (Objective)	Degree Centrality			Relative Betweenness Centrality (%)		
		S1-2	S1	S2	S1-2	S1	S2
14	Supervision and Inspection	0.01	0.015	-	0.019	0.029	-
15	Medical Big Data	0.01	-	0.035	0.019	-	-
16	Early Warning and Emergency Response	0.009	0.013	-	0.019	0.029	-
17	Transportation Support	0.009	0.013	-	0.019	0.029	-
18	Vaccination	0.007	0.009	-	0.019	0.029	-
19	Encouragement and Guidance	0.006	0.008	-	0.019	0.029	-
20	Institutional Improvement	0.006	-	0.018	0.019	-	-
21	Human Resources and Employment	0.004	-	0.013	0.019	-	-
22	Meteorology and Public Health Safety	0.003	0.005	-	-	-	-
23	Compensation and Distribution	0.003	0.004	-	-	-	-
24	Personnel Training/Talent Cultivation	0.001	0.002	-	-	-	-

Table A6. Degree Centrality and Betweenness Centrality of Policy Tools in China's Public Health Policy.

No.	Policy Tool (Instrument)	Degree Centrality			Relative Betweenness Centrality (%)		
		S1-2	S1	S2	S1-2	S1	S2
1	Supervision and Inspection	0.11	0.118	0.098	4.733	3.367	7.165
2	Publicity and Public Opinion	0.106	0.122	0.076	2.574	3.367	2.405
3	Information Release and Feedback	0.105	0.123	0.082	1.37	3.367	1.11
4	Technical Training and Capacity Building	0.071	0.076	0.058	4.733	3.367	4.759
5	Theoretical Knowledge Learning	0.06	0.085	0.018	1.37	3.367	0.198
6	Professional Personnel Training	0.055	0.052	0.061	2.574	3.367	2.405
7	Financial Support	0.047	0.03	0.07	3.334	3.367	3.187
8	Performance Appraisal	0.039	0.008	0.073	0.732	0.152	1.11
9	Maintenance of Hygiene Conditions	0.029	0.051	-	0.22	0.302	-
10	Government Assistance	0.028	0.035	0.025	0.865	0.849	0.238
11	Vaccination	0.026	0.024	0.032	0.865	0.784	0.728
12	Disinfection, Isolation, and Screening	0.024	0.021	0.023	0.22	0.444	-
13	Material Support/Supplies Guarantee	0.024	0.022	0.028	0.865	0.444	0.728
14	Informatization Construction (IT Infrastructure)	0.022	-	0.053	0.732	-	1.11
15	Infrastructure Construction	0.015	0.005	0.025	0.732	0.324	1.11
16	Personal Protection	0.013	0.015	0.01	1.219	1.74	1.11
17	Entry–Exit Quarantine	0.011	0.014	0.007	0.099	0.13	-
18	Research and Development (R&D)	0.011	0.008	0.014	0.865	0.302	1.11
19	International Cooperation	0.01	0.012	0.007	0.423	0.13	0.369
20	Psychological Counseling	0.008	0.013	0.002	-	-	-

https://doi.org/10.3390/systems14050497

Table A6. *Cont.*

No.	Policy Tool (Instrument)	Degree Centrality			Relative Betweenness Centrality (%)		
		S1-2	S1	S2	S1-2	S1	S2
21	Investigation and Research	0.008	0.01	0.006	0.732	0.615	0.198
22	Nucleic Acid Testing (PCR Testing)	0.008	-	0.019	0.54	-	0.728
23	Personnel Mobility Control	0.007	-	0.013	-	-	-
24	Transport Health Quarantine	0.006	0.012	-	0.099	0.13	-
25	Ventilation	0.006	0.008	0.003	0.104	0.13	-
26	Promotion of Resident Health Records	0.006	-	0.016	0.54	-	0.728
27	Simulation Drills	0.005	0.011	-	0.099	0.258	-
28	Establishment of Resident Health Records	0.005	0.001	0.013	-	-	-
29	Strict Punishment in Accordance with Law	0.005	0.001	0.01	0.994	-	2.486
30	Risk Detection	0.004	0.001	0.007	0.37	-	-
31	Pilot Demonstrations	0.004	-	0.014	0.01	-	0.024
32	Reward and Punishment Mechanism	0.003	0.004	-	0.246	0.324	-
33	Restriction of Outdoor Activities	0.003	0.006	-	0.104	0.13	-
34	Resource Sharing	0.003	-	0.01	0.551	0	0.764
35	Entry–Exit Declaration	0.002	0.004	-	-	0.113	0
36	Government Support	0.002	-	0.005	0.896	-	1.267
37	Risk Assessment	0.002	0.003	-	-	-	-
38	Risk Monitoring	0.002	0.001	0.006	0.919	-	0.764
39	Policy Support	0.002	-	0.005	0.945	-	1.346
40	Data Backup	0.002	-	0.005	0.01	-	0.024
41	Standardization/Standard Setting	0.002	-	0.005	0.01	-	0.024
42	Resumption of Work and Production	0.002	-	0.004	1.385	-	1.38
43	Re-examination and Retesting	0.002	-	0.003	-	-	-
44	Follow-up Visits	0.002	-	0.003	-	-	-
45	Suspension of Business Operations	0.002	-	0.003	-	-	-
46	Mask Wearing	0.002	-	0.003	-	-	-
47	Regulated Operation	0.001	-	0.001	-	-	-
48	Contracted Service Projects	0.001	-	0.004	-	-	-
49	Implementation of Guardianship and Care	0.001	-	0.004	-	-	-
50	Physical Exercise	0.001	-	0.002	-	-	-

References

1. Haldane, V.; De Foo, C.; Abdalla, S.M.; Jung, A.-S.; Tan, M.; Wu, S.; Chua, A.; Verma, M.; Shrestha, P.; Singh, S.; et al. Health systems resilience in managing the COVID-19 pandemic: Lessons from 28 countries. *Nat. Med.* **2021**, *27*, 964–980. [CrossRef]
2. Greer, S.L.; King, E.J.; da Fonseca, E.M.; Peralta-Santos, A. The comparative politics of COVID-19: The need to understand government responses. *Glob. Public Health* **2020**, *15*, 1413–1416. [CrossRef]
3. Nan, X.; Iles, I.A.; Yang, B.; Ma, Z. Public health messaging during the COVID-19 pandemic and beyond: Lessons from communication science. *Health Commun.* **2022**, *37*, 1–19. [CrossRef]
4. Van Reisen, M.; Stokmans, M.; Basajja, M.; Ong'ayo, A.O.; Nakazibwe, P.; Kirkpatrick, C. Introduction to the special issue: Data intelligence on patient health records. *Data Intell.* **2022**, *4*, 633–646. [CrossRef]

5. Duckett, J. *The Chinese State's Retreat from Health: Policy and the Politics of Retrenchment*, 1st ed.; Routledge: London, UK, 2011. [CrossRef]
6. Hua, Z.; Yuan, Q. Cross-domain collaboration of digital medical service resources. *J. Syst. Eng.* **2025**, *40*, 738–747. [CrossRef]
7. Zhai, Y.; Guo, R.; Wang, Y. Medical health data sharing strategy based on tripartite evolutionary game. *J. Syst. Eng.* **2025**, *40*, 374–392+420. [CrossRef]
8. Wu, W.; Guo, S. Value consensus, status quo bias and policy change: The case of health policy in China. *J. Public Manag.* **2018**, *15*, 46–57.
9. Fei, T. Building a healthy China: The development process and experience of China's medical and health services under the leadership of the CPC. *Manag. World* **2021**, *37*, 26–40.
10. Zhao, D.; Yang, J. "Dynamic zero-COVID" strategy for controlling the pandemic in Shanghai: Simulation analysis using real data from March to May 2022. *J. Public Manag.* **2023**, *20*, 10–19.
11. Wang, X.; Tang, M.; Yang, X.; Chen, J.; Song, L. Collaborative relationship and evolution features among emergency decision-makers in grave public health emergencies. *Inf. Sci.* **2023**, *41*, 17–25.
12. Head, B.W.; Alford, J. Wicked problems: Implications for public policy and management. *Adm. Soc.* **2015**, *47*, 711–739. [CrossRef]
13. Hsiang, S.; Allen, D.; Annan-Phan, S.; Bell, K.; Bolliger, I.; Chong, T.; Druckenmiller, H.; Huang, L.Y.; Hultgren, A.; Krasovich, E.; et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* **2020**, *584*, 262–267. [CrossRef] [PubMed]
14. Hale, T.; Angrist, N.; Goldszmidt, R.; Kira, B.; Petherick, A.; Phillips, T.; Webster, S.; Cameron-Blake, E.; Hallas, L.; Majumdar, S.; et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat. Hum. Behav.* **2021**, *5*, 529–538. [CrossRef]
15. van Hulst, M.; Metze, T.; Dewulf, A.; de Vries, J.; van Bommel, S.; van Ostaijen, M. Discourse, framing and narrative: Three ways of doing critical, interpretive policy analysis. *Crit. Policy Stud.* **2025**, *19*, 74–96. [CrossRef]
16. de Graaff, S.; Wanzenböck, I.; Frenken, K. The politics of directionality in innovation policy through the lens of policy process frameworks. *Sci. Public Policy* **2025**, *52*, 418–432. [CrossRef]
17. Jiang, Y.; Yu, J. Research on the combined application of policy tools in the management of major public health crises: A textual analysis of the Chinese central government's epidemic control policy. *J. Public Manag.* **2020**, *17*, 1–9.
18. Huang, C.; Yang, C.; Su, J. Identifying core policy instruments based on structural holes: A case study of China's nuclear energy policy. *J. Informetr.* **2021**, *15*, 101145. [CrossRef]
19. Tan, C.; Liang, Y.; Wei, W.; Diao, P.; Chen, X. Text measurement and optimization of China's science and technology personnel evaluation policy based on four-dimensional analytical angles. *Inf. Sci.* **2022**, *40*, 63–71.
20. Zhao, Y.; Wang, Z.; Pei, L. A thematic analysis of public cultural service policy based on policy measurement. *Libr. Inf. Serv.* **2020**, *64*, 66–74.
21. Diana, M.G.; Mascia, M.L.; Tomczyk, Ł.; Penna, M.P. The Digital Divide and the Elderly: How Urban and Rural Realities Shape Well-Being and Social Inclusion in the Sardinian Context. *Sustainability* **2025**, *17*, 1718. [CrossRef]
22. Sun, H.; Liang, Y.; Yuan, T.; Wang, Y. Content analysis and market impact of PPP policy. *J. Syst. Eng.* **2024**, *39*, 175–188. [CrossRef]
23. Ge, T.; Chen, X.; Geng, Y.; Yang, K. Does regional collaborative governance reduce air pollution? Quasi-experimental evidence from China. *J. Clean. Prod.* **2023**, *419*, 138283. [CrossRef]
24. Li, Y.; Yang, R.; Lu, Y. A privacy risk identification framework of open government data: A mixed-method study in China. *Gov. Inf. Q.* **2024**, *41*, 101916. [CrossRef]
25. Inaam ul haq, M.; Li, Q. Revealing the trends in the academic landscape of the health care system using contextual topic modeling. *Data Intell.* **2023**, *5*, 923–946. [CrossRef]
26. Gu, D.; Liu, H.; Zhao, H.; Yang, X.; Li, M.; Liang, C. A deep learning and clustering-based topic consistency modeling framework for matching health information supply and demand. *J. Assoc. Inf. Sci. Technol.* **2024**, *75*, 174–192. [CrossRef]
27. Miao, H.; Quan, Q.; Shu, X. Topic identification and quantitative analysis of science and technology talent policy texts based on BERTopic: A case study of three northeast provinces in China. *J. Mod. Inf.* **2025**, *45*, 110–121.
28. Gu, D.; Wang, Q.; Chai, Y.; Yang, X.; Zhao, W.; Li, M.; Zolotarev, O.; Xu, Z.; Zhang, G. Identifying the risk factors of allergic rhinitis based on Zhihu comment data using a topic-enhanced word-embedding model: Mixed method study and cluster analysis. *J. Med. Internet Res.* **2024**, *26*, e48324. [CrossRef]
29. Tao, C.; Wu, J.; Shui, D.; Yu, Y. Orientation and trend: Topic mining and evolution of data element trading policy. *Inf. Stud. Theory Appl.* **2024**, *47*, 39–48.
30. Zhang, Q.; Li, R.; Wu, S.; Liu, X.; Wang, H. COKG-QA: Multi-hop question answering over COVID-19 knowledge graphs. *Data Intell.* **2022**, *4*, 472–493. [CrossRef]
31. Cen, R.; Zhou, L. Analysis on the evolution of standardization policy in the field of digital economy. *Sci. Manag. Res.* **2022**, *40*, 105–115.
32. Zhao, Z.; Li, Y.; Sheng, S.; Qu, A. Evolutionary trajectory and influencing factors of regional carbon emission reduction policy collaboration: A network analysis perspective. *Public Adm. Policy Rev.* **2024**, *13*, 20–33.

33. Chen, S.; Miao, C.; Zhang, Q. Understanding the evolution of China's green shipping policies: Evidence from social network analysis. *J. Clean. Prod.* **2024**, *482*, 144204. [CrossRef]
34. Yao, H.; Zhang, C. A bibliometric study of China's resource recycling industry policies: 1978–2016. *Resour. Conserv. Recycl.* **2018**, *134*, 80–90. [CrossRef]
35. Liu, M.; Yuan, S.; Li, B.; Zhang, Y.; Liu, J.; Guan, C.; Chen, Q.; Ruan, J.; Xie, L. Chinese Public Attitudes and Opinions on Health Policies During Public Health Emergencies: Sentiment and Topic Analysis. *J. Med. Internet Res.* **2024**, *26*, e58518. [CrossRef]
36. Walker, C.; Moulis, A. Understanding policy transfer through social network analysis: Expanding methodologies with an intensive case study approach. *Policy Sci.* **2022**, *55*, 693–713. [CrossRef]
37. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 3982–3992. [CrossRef]
38. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794. [CrossRef]
39. Wasserman, S.; Faust, K. *Social Network Analysis: Methods and Applications*; Cambridge University Press: Cambridge, UK, 1994. [CrossRef]
40. Provan, K.G.; Kenis, P. Modes of Network Governance: Structure, Management, and Effectiveness. *J. Public Adm. Res. Theory* **2008**, *18*, 229–252. [CrossRef]
41. Meng, Q. Strengthening public health systems in China. *Lancet Public Health* **2022**, *7*, e987–e988. [CrossRef]
42. Gong, Y.; Yang, X. Understanding strategies for digital government transformation: A strategic action fields perspective. *Int. J. Inf. Manag.* **2024**, *76*, 102766. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

systems

MDPI

Article

The Impact of Digital Government on Regional Scientific and Technological Innovation Capacity

Zhengang Zhang [1,2] and Defei Wang [2,3,*]

[1] School of Business Administration, South China University of Technology, Guangzhou 510640, China
[2] Guangzhou Digital Innovation Research Center, Guangzhou 510640, China
[3] School of Public Administration, South China University of Technology, Guangzhou 510640, China
* Correspondence: 202420139126@mail.scut.edu.cn

Abstract

Background and Purpose: Contemporary digital government initiatives in China face a well-documented real-world paradox: massive fiscal outlays on digital governance coexist with marked disequilibrium in regional innovation returns. Two structural mismatches define this paradox. First, local governments overwhelmingly prioritize high-visibility hardware investments such as data centers and large AI models, while neglecting deep-seated institutional reforms including cross-departmental business process reengineering and factor market liberalization. The pervasive phenomenon of "aggregated but non-interoperable data, and interoperable data left unused" reflects a severe asynchrony between rapid technological deployment and lagging institutional restructuring. Second, comparable digital investments yield vastly divergent innovation dividends across eastern, central, and western regions, with regional divergence entrenching into a rigid "higher in the east lower in the west, higher in the south lower in the north" pattern. Extant literature, largely confined to the lens of "instrumental rationality," reduces digital government to an exogenous technological variable, leaving it unable to explain this core practical puzzle of "homogeneous inputs generating heterogeneous returns." Moving beyond the narrow "technology-enabled governance" narrative, this study draws on the Digital-Era Governance (DEG) paradigm to investigate the actual impact of institutional restructuring on regional scientific and technological innovation capacity, aiming to provide empirical evidence to unlock the inefficiency lock-in prevalent in digital governance practices. **Research Methods:** This study uses 280 prefecture-level cities and above in China from 2018 to 2023 as the research sample and constructs a two-way fixed-effects model for benchmark regression analysis. To address endogeneity, the average level of digital government development in other cities within the same province is used as an instrumental variable, and the 2SLS method is employed to identify the causal effect. On this basis, a series of robustness checks are conducted, including excluding the special impact of the COVID-19 pandemic, substituting core variable measures, and decomposing the dimensions of the core explanatory variables, to ensure the reliability of the research conclusions. For mechanism identification, the Bootstrap sampling method is used to test the dual mediating effects of "digital industry agglomeration" and "resource misallocation alleviation"; furthermore, moderating effects and heterogeneity analysis models are introduced to reveal the boundary constraints of regional economic development levels and city types on the empowerment effect. **Main Findings:** Empirical results show that: (1) Digital government construction significantly improves regional scientific and technological innovation capacity, and this conclusion remains valid after endogeneity treatment and robustness checks. (2) Mechanism analysis demonstrates that digital government drives innovation through the dual paths of "promoting digital industry agglomeration"

and "alleviating resource misallocation," with the marginal contribution of alleviating resource misallocation being significantly higher than that of industrial agglomeration. This suggests that, in transitional economies, eliminating institutional frictions in factor mobility brings greater innovation dividends than simply building physical spatial clusters. (3) Moderating effects indicate that the higher the level of regional economic development, the stronger the innovation empowerment effect of digital government. (4) Heterogeneity analysis further reveals that the innovation dividends of digital government are significant only in non-resource-based cities, non-central cities, and large and medium-sized cities, while in resource-based cities, central cities, and small cities, the effects are systematically absorbed and not significant. **Research Conclusions and Contributions:** This study breaks through the ontological limitations of existing research that views digital government as a technological tool, grounding it within the DEG theoretical framework and confirming that digital government is an institutional force in the reconstruction of regional innovation ecosystems. The findings suggest to policymakers that digital government construction should promote a shift from a "technology-oriented" to an "institution-oriented" approach. The policy focus should shift from mere infrastructure expansion to the elimination of deep-seated institutional frictions, the improvement of factor allocation efficiency, and the advancement of gradients and the implementation of classified governance, all guided by regional economic foundations and heterogeneity characteristics.

Keywords: digital government; regional scientific and technological innovation capacity; digital industry agglomeration; resource misallocation; regional economic development level

1. Introduction

Amid intensifying global tech rivalries and the prevailing pursuit of high-quality development, regional scientific and technological innovation capacity has become the decisive cornerstone for economies to build lasting competitive advantages [1,2]. Its generative mechanisms are rooted not only in the accumulation of R&D capital and human capital, but are increasingly constrained by deep-seated institutional environments and public governance paradigms [3,4]. Empirical observation further reveals that linear growth models relying solely on factor inputs can no longer sustain rising innovation vitality, with growing asymmetry between investment and returns becoming increasingly prominent.

Confronting industrial restructuring triggered by the new wave of technological revolution, countries worldwide are racing to position digital government construction as a pivotal strategic fulcrum for reshaping national innovation systems [5,6]. China has likewise integrated its dual national strategies of "Digital China" and "Innovation-Driven Development" into a cohesive policy framework, with local governments scaling up annual investment in digital government initiatives. Industry estimates indicate national annual spending on digital government surged from roughly RMB 180 billion in 2015 to nearly RMB 600 billion in 2024, with cumulative investment exceeding RMB 4 trillion over the decade and a compound annual growth rate of 10–15%. Yet field audits expose glaring implementation biases: local governments overwhelmingly prioritize high-visibility hardware assets such as data platforms, government cloud systems, and large language models, while progress on deep-seated institutional reforms—including cross-departmental business process reengineering, overhaul of administrative approval logic, and factor market liberalization—lags significantly. The pervasive phenomenon of "aggregated but non-interoperable data, and interoperable data left unused" persists, and entrenched bureaucratic inertia driven by siloed vertical (tiao) and territorial (kuai) admin-

istrative logics remains largely unshaken [7,8], severely constraining digital government's substantive enabling effect on regional innovation ecosystems.

While extant empirical studies broadly corroborate the positive enabling effect of digital government on regional innovation [9], spatial heterogeneity in innovation outputs is strikingly pronounced [10]. Official assessments confirm a rigid "higher in the east lower in the west, higher in the south lower in the north" spatial pattern in local government digital service capacity, with gaps continuing to widen rather than converge: identical digital project investments yield up to three times more patent output in eastern coastal cities than in inland counterparts. This regional heterogeneity and opacity of transmission mechanisms expose a core theoretical lag in the current literature—existing research lacks systematic analysis of the specific governance logic through which digital government reconstructs regional innovation systems, and the underlying drivers of efficacy divergence. More critically, this theoretical void has trapped local policy practice in a "high-input, low-return" dilemma: many less-developed regions ignore their local factor endowments, pouring substantial digital budgets into vanity projects such as smart city dashboards, while grassroots officials prioritize building "digital bonsai" (superficial showcase projects) to meet assessment metrics, focusing on surface-level compliance over substantive enabling, resulting in massive waste of fiscal resources.

Existing research explores the mechanisms linking digital government to regional scientific and technological innovation from multiple perspectives, yet shares a latent ontological limitation: it tends to treat digital government as an exogenous "instrumental variable" or "environmental variable" detached from the innovation system, rather than an endogenous "governance paradigm shift" in the process of institutional change. Specifically, Transaction Cost Economics presumes the core function of institutions is to reduce information frictions and contract enforcement costs; accordingly, extant literature deconstructs digital government as a governance tool that lowers transaction costs in innovation activities via process standardization and information sharing [11,12]. In reality, however, even after massive investment in cross-departmental shared platforms, the fragmented incentive structure of hierarchical organizations remains unchanged, and intrinsic motivations for data hoarding persist, rendering claims of reduced transaction costs hollow. The Resource-Based View frames digital government as the supply side of public data resources, emphasizing how heterogeneous data assets enhance a region's capacity to attract innovation factors [13]. Yet relevant data shows over 60% of public datasets remain inaccessible to innovators due to opaque access protocols and restrictive licensing requirements, creating obvious bottlenecks in releasing the value of these resources. New Structural Economics incorporates digital government into its portfolio of "facilitative" policy tools, advocating for aligning digital infrastructure investment with local factor endowments to amplify innovation dividends [14,15]. In practice, however, many localities ignore factor matching principles, blindly replicating digital projects from developed regions and causing misallocation of resources. Institutional Theory views digital government as a source of coercive or normative pressure, driving local governments to adopt isomorphic digital governance practices to gain legitimacy [16,17]. This pressure has directly spawned a wave of performative digitalization: local officials prioritize meeting superficial compliance metrics (e.g., number of online services) over delivering substantive institutional innovations, creating "digital Potemkin projects" that mask underlying governance shortcomings. The Technology–Organization–Environment (TOE) framework disassembles digital government into a three-dimensional combination of technological characteristics, organizational readiness, and environmental conditions; scholars such as Zhu have applied this framework to analyze interactive effects of technology, organization, and environment on the assimilation of digital innovation diffusion [18,19]. Yet its

static design fails to capture the dynamic interplay between technology and institutions: it treats digital government as a fixed exogenous input rather than an evolving governance paradigm, naturally unable to explain why identical TOE configurations yield drastically different innovation returns across regions.

While the aforementioned perspectives offer valuable insights into specific facets of digital government's impact, they remain constrained by the shackles of "instrumental rationality" and fail to capture its dynamic nature as a paradigmatic shift in Digital-Era Governance (DEG) [20]. Whether framing digital government as a cost-reduction tool, resource supplier, factor-matching instrument, or institutional pressure source, these views overlook its core function of dismantling inherent bureaucratic fragmentation via Reintegration, and its capacity to restructure deep-seated interaction structures between government and innovators via Needs-based Holism. This excessive focus on instrumental attributes leaves extant literature ill-equipped to explain the real-world paradox of "identical inputs yielding heterogeneous returns"—precisely the core proposition of "institutional reconstruction" that DEG theory seeks to address.

To systematically bridge this gap, this study leverages a panel dataset covering 280 Chinese prefecture-level and above cities from 2018 to 2023, marking the first empirical effort to introduce DEG theory into regional innovation research and construct an integrated analytical framework spanning main effects, dual mediation pathways, and moderating mechanisms. Grounded in DEG's triple logic, digital government is not a simple overlay of information technologies, but reconstructs the institutional environment of regional innovation systems via four pathways: cross-departmental data sharing, innovation-demand-responsive service delivery, automated administrative workflows, and open public data access. Accordingly, we derive and empirically test three core hypotheses: H1 posits a significant positive main effect of DEG-informed digital government construction on regional scientific and technological innovation capacity. H2a and H2b further posit that this effect is transmitted through two parallel mediation channels: alleviation of resource misallocation (the factor market correction function of Reintegration) and promotion of digital industrial agglomeration (the industrial ecosystem induction effect of Needs-based Holism). H3 posits that regional economic development level positively moderates the relationship between digital government and regional innovation, such that the enabling effect of DEG strengthens with higher levels of regional economic development.

The marginal contributions of this study are threefold. First, reframing of the theoretical lens. We transcend the ontological limitation of treating digital government as an exogenous instrumental variable, systematically introducing DEG theory into regional innovation research to achieve a cognitive leap from "instrumental rationality" to "paradigm shift," providing a novel theoretical lens for understanding the nexus between digital governance and innovation. Second, empirical deconstruction of transmission mechanisms. We unpack the "black box" linking digital government to regional innovation, empirically identifying two parallel and distinct mediation channels—"alleviation of resource misallocation" and "digital industrial agglomeration"—which clarifies the ambiguities of vague attributions such as "business environment improvement" in prior research, and directly addresses the practical confusion of local governments facing "investment with no visible returns." Third, contextualized delineation of boundary conditions. We explicitly identify the moderating role of regional economic development in conditioning digital governme's innovation-enabling effect, alongside its heterogeneous characteristics across regions and city types. This provides precise empirical evidence to resolve the "high-input, low-return" dilemma in local digital governance, preventing less-developed regions from blindly copying digital construction models of developed areas and reducing wasteful misallocation of fiscal resources.

2. Literature Review and Research Hypotheses

2.1. The Relationship Between Digital Government and Regional Scientific and Technological Innovation Capacity

Digital government, as a strategic pivot in reshaping national innovation systems, has garnered widespread consensus regarding its centrality within academic scholarship. While extant literature broadly corroborates the positive enabling effect of digital government on regional innovation, a growing theoretical fault line has emerged with ongoing policy implementation: existing studies struggle to account for the disequilibrium between large-scale digital government investments and heterogeneous regional innovation returns. This disconnect stems not from scattered explanatory gaps, but from three systemic limitations endemic to prevailing research paradigms.

First, theoretical lag in addressing technology-institution asynchrony. Literature anchored in Transaction Cost Economics (TCE) predominantly operationalizes digital government as a plug-and-play tool for reducing information frictions [11,21] implicitly assuming a linear causality wherein technological deployment automatically triggers leaps in governance efficacy. Empirical evidence, however, suggests that the institutional rigidity of hierarchical structures often neutralizes technological dividends, precipitating "synergy bottlenecks" in data and business operations [8]. The TCE framework, lacking theoretical tools to address "institutional restructuring lags," persistently fails to explain why substantial hardware investments fail to translate into substantive innovation enablement. Furthermore, while Layne and Lee proposed a seminal four-stage model for achieving "fully functional" e-government—moving from catalog publishing to online transactions and finally vertical/horizontal integration—it remains largely descriptive of technical maturity. It fails to theorize the deep-seated institutional fragmentation that impedes the transition from mere transactional efficiency to systemic governance innovation, leaving a vacuum in explaining why many localities stagnate at the "online service delivery" stage despite heavy investment [22].

Second, presumptive bias towards homogeneous variables in regional divergence research. Although extensive empirical work documents spatial heterogeneity in digital government's innovation effects [21], the dominant TOE framework continues to treat digital government as a homogeneous independent variable [18]. This treatment obscures fundamental disparities in governance paradigm maturity across regions—while some jurisdictions have completed the paradigm shift towards Digital-Era Governance (DEG), others remain mired in the nascent stages of e-government. Public administration scholarship has systematically critiqued the TOE framework for its "black-boxed" technological dimensions, neglect of micro-mechanisms, and reliance on static causal chains. Consequently, it proves particularly anemic in explaining phenomena where "identical inputs yield heterogeneous returns."

Third, ontological constraints of instrumental rationality in mainstream perspectives. While the Resource-Based View, Institutional Isomorphism Theory, and Structural Economics characterize digital government functionality across various dimensions, they converge on a shared premise: digital government is treated as a technological variable or environmental factor exogenous to the innovation system. This instrumental lens fixates on marginal cost-reduction effects, wholly overlooking digital government's endogenous institutional attributes as a governance paradigm shift. DEG theory, as articulated by Dunleavy et al. [20], fundamentally corrects this by positing that digital government is not a mere technological overlay but a reconstruction of governance itself through three logics: Reintegration, Needs-based Holism, and Digitization. By clinging to instrumental rationality, extant literature remains incapable of unpacking the "similar input—heterogeneous

return" black box or elucidating the specific institutional pathways through which digital government influences innovation.

Collectively, these limitations point to a critical theoretical lacuna: to provide a valid analytical lens for regional innovation system (RIS) institutional analysis, digital government must be reframed from an "exogenous tool variable" to an "endogenous governance paradigm." Traditional innovation research, rooted in the Knowledge Production Function [1], attributes innovative output to R&D capital and high-skilled labor. Furthermore, the RIS framework advanced by Cooke and Asheim & Isaksen emphasizes that innovation capacity is an emergent collective property arising from interactive learning among diverse actors within a specific institutional milieu [23,24]. However, RIS literature has long marginalized "digitalization," relegating it to a control variable or mere technological backdrop, failing to address how a paradigm shift in governance reconstructs the factor allocation rules and interaction topology of RIS. DEG theory bridges this gap precisely by aligning Reintegration with the dissolution of bureaucratic fragmentation, Needs-based Holism with the institutionalization of service logic, and Digitization with the friction-reduction in the government–innovator interface. It thus constitutes an analytically indispensable framework for bridging the aforementioned theoretical gaps, rather than a mere artifact of theoretical construction. Crucially, as Adner and Kapoor demonstrate in the context of technological interdependence, the structure of an ecosystem dictates how value is created and captured. By reconfiguring the "interaction topology" between government, firms, and research institutions, DEG transforms the innovation ecosystem from a loosely coupled network into a cohesive structure capable of navigating the complexities of new technology generations, thereby providing the missing institutional micro-foundation for RIS in the digital era [6,20].

Drawing on Digital-Era Governance (DEG) theory, this study posits that digital government is not merely a cumulative layering of technological applications, but reconstructs the institutional bedrock of regional innovation systems through three core mechanisms.

First, the Reintegration mechanism dismantles the siloed structure of traditional bureaucracy. By enabling cross-departmental data sharing and business collaboration, digital government eliminates "data islands" and redundant approval barriers that impede the flow of innovation factors. This significantly reduces the ex-ante search costs and ex-post compliance execution costs faced by innovators [21], allowing them to allocate more resources to core R&D activities.

Second, the Needs-based Holism mechanism reconfigures the logic of service delivery. Digital government shifts the unit of service provision from "departmental functions" to "innovator demands." Through one-stop service platforms and standardized open data interfaces, it accurately aligns with firms' demands for policies, funding, and knowledge information, thereby mitigating innovation decision biases and search costs caused by information asymmetry.

Finally, the Digitization mechanism reduces systemic coordination costs. The automation and platformization of government affairs reconstruct the interaction interface between government and innovators. This lowers the asset specificity and coordination resistance of cross-organizational and cross-regional collaboration, facilitating the recombination of talent, technology, and capital in a low-friction environment. Consequently, it provides institutional safeguards for long-cycle basic research and frontier technological innovation [25,26].

Through these mechanisms, digital government optimizes the operational efficiency and the quality of factor matching in regional innovation ecosystems. Empirical evidence further indicates that digital government construction significantly increases regional

innovation investment and enhances patent output quality [26]. Accordingly, we propose the following hypotheses:

H1: *The development of digital government significantly and positively promotes regional technological innovation capability.*

2.2. The Mediating Role of Digital Industry Agglomeration

The development of the digital economy necessitates a shift in regional industrial structure from isolated interfirm competition to ecosystem-level synergy. Rooted in the regional innovation system (RIS) theory, industrial agglomeration serves as a critical spatial and organizational carrier for regional technological innovation [27]. Recent scholarship further confirms that this logic extends to the digital industry context—the geographical and organizational concentration of digital firms and related institutions can generate analogous agglomeration dynamics for regional technological progress through shared infrastructure, specialized labor pools, and high-frequency knowledge spillovers [28,29]. However, the extant literature lacks sufficient theoretical explication of the institutional logic driving the formation of such high-quality digital industry agglomeration. Mainstream research predominantly attributes agglomeration to factor endowments and market mechanisms, treating it as a spontaneous outcome of firm location choices guided by factor prices. This perspective overlooks the uniqueness of digital industries: unlike traditional manufacturing, which relies on static resource accumulation, digital industries exhibit strong network effects and positive externalities, making their agglomeration process highly dependent on dynamic institutional arrangements to coordinate complex cross-boundary collaborations [30]. Without clarifying how digital government, acting as an "architect of governance architecture," induces agglomeration through institutional restructuring, it remains difficult to explain why similar digital investments yield drastically divergent innovation outcomes [31].

Digital-Era Governance (DEG) theory provides a precise framework for analyzing this institutional induction mechanism. First, the Reintegration mechanism addresses the institutional fragmentation constraining agglomeration. Advanced digital industries (e.g., artificial intelligence, industrial internet) require seamless interoperability across industries and departments. Data silos and conflicting regulatory rules under traditional bureaucracy significantly increase transaction costs for industrial chain collaboration. Digital government breaks down administrative barriers to factor mobility through mandatory cross-department data sharing and unified digital platform construction [30]. This institutional-level "rewiring" not only reduces firms' compliance costs but also creates frictionless institutional interfaces for the geographical concentration of digital industry chains, rendering high-density industrial collaboration economically viable.

Second, the Needs-based Holism mechanism enables digital government to reorganize public service delivery logic according to the genuine needs of digital innovators. Unlike traditional industrial clusters reliant on generic preferential policies (e.g., land, tax breaks), digital industry agglomeration exhibits high dependence on specific production factors such as public computing power, open data, and testing scenarios [29,31]. By shifting the unit of service delivery from "departmental functions" to "innovation demands," digital government integrates and opens public computing resources and desensitizes government datasets on demand, reducing transaction costs for digital firms to search, negotiate, and access critical production factors. This demand-centric institutional orchestration—rather than mere hardware deployment—compresses firms' sunk costs and entry barriers, proactively guiding and accelerating the formation of high-quality digital industry agglomeration instead of passively awaiting slow market evolution.

Finally, the Digitization mechanism strengthens the internal network effects within agglomerations. The automation and platformization of government affairs reconstruct the interface for interaction between government and innovators, transforming offline, discrete regulation and services into an online, continuous ecosystem of governance. This digital governance environment enhances the intensity of connections among firms, research institutions, and innovators within the agglomeration zone, accelerating the spillover of tacit knowledge and the recombination of explicit knowledge, thereby transforming physical spatial agglomeration into functional innovation agglomeration.

Consequently, digital industry agglomeration induced by DEG mechanisms is no longer merely the geographical clustering of firms but evolves into an innovation ecosystem characterized by high connectivity and high coordination efficiency. Through scale effects, knowledge recombination, and network synergy, such systems directly enhance regional scientific and technological innovation capacity [8,30]. Admittedly, prior research has demonstrated that digital government improves information accessibility and the external environment for innovation by enhancing transparency, opening data, and supplying universal digital infrastructure, However, these pathways essentially position digital government at the level of information provision and technological deployment, failing to capture its impact as a DEG paradigm shift on the structural reorganization of innovation ecosystems. As an independent mediating mechanism, digital industry agglomeration reveals how DEG translates institutional advantages into spatial organizational advantages—a function irreplaceable by mere technological supply or environmental optimization. Based on the foregoing theoretical deductions, this paper proposes the following hypotheses:

H2a: *Digital government enhances regional scientific and technological innovation capacity by promoting digital industry agglomeration.*

2.3. The Mediating Role of Resource Misallocation

In idealized markets, factors such as talent, capital, and technology spontaneously achieve efficient allocation through price signals [32]. In reality, however, information asymmetry, externalities, and institutional frictions are pervasive, preventing factors from flowing toward innovation activities with the highest marginal output and severely constraining regional scientific and technological innovation capacity [4]. In transitional economies like China, institutional distortions—such as market segmentation driven by local protectionism, credit discrimination favoring specific ownership types, and institutional barriers to labor mobility—further exacerbate resource misallocation in factor markets [33].

Such distortions not only hinder total factor productivity growth but also constitute a systemic bottleneck for innovation activities [34]. Simultaneously, by unifying regulatory enforcement standards and tax collection rules across regions, it eliminates market segmentation caused by local protectionism. This enables innovation factors to recombine freely based on marginal output efficiency rather than administrative boundaries, significantly reducing institutional rents derived from resource misallocation.

Second, the Needs-based Holism logic targets the elimination of policy-induced factor price distortions. The traditionally fragmented regulatory architecture often leads to the dissipation of policy tools and implementation deviations, manifesting as the haphazard allocation of subsidies and distorted entry costs imposed by multi-headed approval processes and hidden licensing barriers [35]. Digital government shifts the logic of service delivery from "departmental function supply" to "innovation demand response." By utilizing big data profiling to implement data-driven, targeted industrial policy support, it effectively compresses rent-seeking space and builds a fairer competitive environment by reducing information asymmetry and discretionary approval bias. This institutional orchestration

ensures that various innovation entities can access critical resources based on efficiency rather than rent-seeking capability, thereby correcting price signal distortions in factor markets at their source.

Third, the Digitization logic strengthens the dynamic optimization capacity of factor allocation. The automation and platformization of government affairs reconstruct the interaction interface between government and innovators; leveraging real-time data monitoring and algorithmic matching, factors are redirected from low-efficiency "zombie firms" and saturated sectors to high-productivity innovation activities such as green technology and advanced manufacturing. This raises the marginal return to innovation investment [36]. This dynamic allocation mechanism remedies the lag of traditional static factor allocation models and unlocks innovation potential previously trapped in inefficient sectors.

While extant studies suggest digital government can promote innovation by enhancing transparency or strengthening intellectual property protection, these pathways fail to address the underlying operational rules of factor markets. As an independent mediating mechanism, the alleviation of resource misallocation uniquely characterizes how DEG translates institutional advantages into factor allocation efficiency—a function irreplaceable by mere environmental optimization. Based on this theoretical deduction, we propose the following hypothesis:

H2b: *Digital government enhances regional scientific and technological innovation capacity by alleviating resource misallocation.*

2.4. The Moderating Role of Regional Economic Development Level

Although the preceding sections have theoretically demonstrated, based on Digital-Era Governance (DEG) theory, that digital government enhances regional scientific and technological innovation capacity through Reintegration, Needs-based Holism, and Digitization, this enabling effect varies significantly across regions. The efficacy of institutional arrangements is highly contingent upon the macroeconomic fundamentals within which they are embedded [37]. This implies that identical digital governance models may yield drastically divergent innovation returns under different economic conditions. This observation prompts a deeper investigation into the boundary role of regional economic development levels in this process.

Drawing on DEG theory, regional economic development levels moderate the strength of digital government's innovation-enabling effect primarily by influencing the operational scope of its three core mechanisms.

First, the effective functioning of the Digitization mechanism is highly dependent on the completeness of regional digital infrastructure. In economically developed regions, robust fiscal capacity supports the widespread deployment of new infrastructure such as 5G networks, data centers, and industrial internet platforms, providing the physical carriers necessary for the advanced functionalities of digital government [38]. This infrastructural completeness ensures that institutional dividends can bridge the "digital divide" and reach a wide spectrum of innovators [37]. Conversely, in regions with lower levels of economic development, limited fiscal resources are often prioritized for guaranteeing basic public services, leading to insufficient investment in new digital infrastructure. Consequently, the Digitization mechanism faces constraints from a weak "hardware foundation," capping its innovation-enabling potential.

Second, the effectiveness of the Needs-based Holism mechanism is constrained by the absorptive capacity and conversion efficiency of innovation actors. Economically developed regions typically host high-skilled human capital, R&D-intensive firms, and sophisticated financial markets. These actors exhibit higher elasticity of demand for digital government services, open data resources, and innovation policies [39]. Grounded in absorptive capacity

theory, these entities are better positioned to rapidly translate the transaction cost reductions facilitated by digital government into substantive R&D outputs [40]. Conversely, in less-developed regions, innovation actors (particularly SMEs) are often constrained by limited technological reserves and financial resources, resulting in insufficient responsiveness to the institutional conveniences offered by the Holism mechanism. This "capability gap" diminishes the actual impact of digital government on innovation.

Finally, the Reintegration mechanism requires synergy with the maturity of the regional institutional environment. Developed regions typically possess more sophisticated market-oriented, rule-of-law, and industrial ecosystems. Such environments resonate positively with the transparent, efficient, and rules-based governance philosophy advocated by digital government. Empirical evidence confirms that the innovation-promoting effect of digital government is more pronounced in regions with superior institutional environments [36]. This is because a sound institutional environment reduces friction and distortion in policy implementation, generating more positive market feedback to the efforts of the Reintegration mechanism in breaking down administrative barriers and promoting the free flow of factors [4,41].

In summary, regional economic development levels systematically moderate the intensity of digital government's innovation enablement by influencing the operational space of the three core DEG mechanisms. Based on this theoretical deduction, we propose the following hypothesis:

H3: *In regions with a higher level of regional economic development, digital government exhibits a stronger effect in improving regional scientific and technological innovation capacity.*

The theoretical model of this paper is shown in Figure 1:

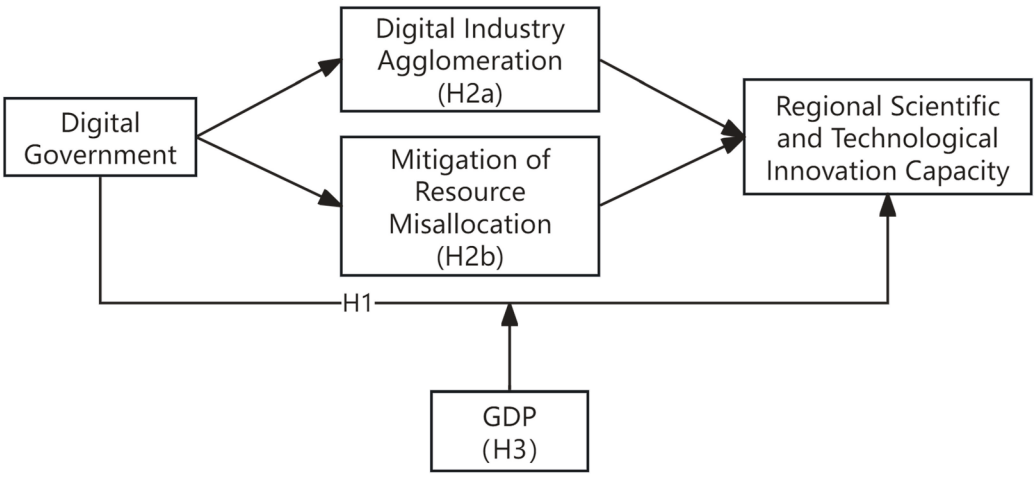


Figure 1. Theoretical model.

3. Research Design

3.1. Variable Definition and Data Description

3.1.1. Dependent Variable

Regional scientific and technology innovation capacity (Tech): The dependent variable in this study is regional scientific and technological innovation capacity, which serves as a core dimension for evaluating regional innovation development. Existing studies commonly quantify this capacity using innovation output indicators, with typical proxies including patent applications, patent grants, and awards for scientific and technological achievements [8,42]. Compared with patent applications (which carry the risk of non-grant) and science and technology awards (which lack a unified official statistical caliber), patent

grants can more accurately reflect the effective outcomes and technological commercialization capacity of regional innovation activities. Therefore, to ensure indicator accuracy and robustness, this study adopts the number of patent grants as the proxy variable for regional scientific and technological innovation capacity.

3.1.2. Core Independent Variable

Digital government (DG): The core independent variable is digital government. Drawing on authoritative evaluation systems in the digital government field, this study follows the measurement framework from the Report on the Development of Internet Service Capability of Local Governments in China by the University of Electronic Science and Technology of China. We measure regional digital government development comprehensively across three dimensions: service supply capability, service responsiveness, and service intelligence. This indicator system reflects the comprehensive ability of local governments to use digital technologies to optimize government services and improve governance effectiveness from the perspective of digital performance, and has been widely adopted in high-quality academic research. The specific indicator system is detailed in Table 1.

Table 1. Evaluation index system for digital government capabilities.

First-Level Indicator	Second-Level Indicator	Third-Level Indicator
Service Provision Capability (40%)	Catalog Coverage Capability (30%)	Responsibility List (25%) Power List (25%) Government Information Disclosure Directory (25%) Public Service List (25%)
	Application Integration Capability (30%)	Platform Integration Capability (40%) Platform Application Capability (45%) Data Openness (15%)
	Service Accessibility Capability (40%)	Social Security Field (12%) Education Field (8%) Medical Field (12%) Employment Field (8%) Housing Field (8%) Transportation Field (8%) Enterprise Establishment & Change (16%) Enterprise Operation and Tax Payment (14%) Innovation and Entrepreneurship Field (14%)
Service Responsiveness Capability (40%)	Service Request Acceptance Capability (30%)	Interactive Complaint Handling Capability (50%) Service Request Acceptance Capability (50%)
	Service Request Response Capability (35%)	First-Class Standard for Online Government Service Handling (20%) Second-Class Standard for Online Government Service Handling (20%) Third-Class Standard for Online Government Service Handling (30%) Fourth-Class Standard for Online Government Service Handling (30%)

Table 1. *Cont.*

First-Level Indicator	Second-Level Indicator	Third-Level Indicator
Service Intelligence Capability (20%)	Interactive Request Feedback Capability (35%)	Request Response Capability (40%) Request Application Capability (30%) Proactive Perception Capability (30%)
	Application Adaptation Capability (40%)	Functional Adaptability (65%) Application Expansion (35%)
	Intelligent Interaction Capability (40%)	Intelligent Search Capability (50%) Intelligent Q&A Capability (50%)
	Personalized Service Capability (20%)	Customized Service (75%) Intelligent Push (35%)

Source: Author's compilation based on Report on the Development of Internet Service Capability of Local Governments in China (2019), Institute of Smart Governance, University of Electronic Science and Technology of China/Social Sciences Academic Press [43].

3.1.3. Mediating Variables

Digital industry agglomeration (digagg): In accordance with the Statistical Classification of the Digital Economy and its Core Industries (2021), the digital industry includes digital equipment manufacturing, digital information transmission, digital technology services, digital content and media, and related services. Based on the measurement framework for digital industry agglomeration proposed by Liu et al., combined with annual city-level data on digital industry employment and total urban employment, this study uses the location entropy index to calculate the agglomeration level of digital industries at the city level [44]. This method quantifies the relative concentration of digital industries in a given region compared with the overall employment structure, effectively eliminating the interference of city size differences on agglomeration characteristics.

Resource misallocation (misall): Based on the classic theory of factor allocation distortion developed by Hsieh and Klenow [31], as well as the composite index framework proposed by Chinese scholar Bai Junhong and his co-authors [33], this study adopts the composite method of relative gaps in factor marginal outputs to calculate the city-level resource misallocation index. A larger index value indicates a higher degree of distortion in the allocation of capital and labor, and an index value of 0 means that factors achieve a fully optimal allocation. All data for the calculation are sourced from official and authoritative statistical materials such as the China City Statistical Yearbook and China Fixed Asset Investment Statistical Yearbook, which guarantee the objectivity, continuity, and comparability of the estimation results.

3.1.4. Moderating Variable

Regional economic development level (GDP): Following relevant studies, this study uses the per capita gross regional product (logarithm) to measure the regional economic development level. This indicator effectively eliminates the influence of city size differences and more accurately reflects regional economic development, and has been widely used in studies on regional innovation and digital-related topics [35].

3.1.5. Control Variables

With reference to variable settings in the literature, this study introduces human capital level (human) [45], foreign direct investment level (open) [46] and industrial structure (IS) [47] as control variables, measured by the logarithm of the number of students enrolled in regular higher education institutions, the ratio of actual foreign direct investment to urban GDP, and the ratio of value-added of the tertiary industry to urban GDP, respectively. Meanwhile, considering that the rapid development of digital technology is continuously

changing the external environment for regional innovation, this study also incorporates the financial development level (Financial) into the control variable system. Innovation activities in the digital sector are typically characterized by high investment, high risk, and high return, and most innovation activities rely heavily on external financial support, where financial institutions play a key role in capital supply. Therefore, this study uses the logarithm of the year-end loan balance of financial institutions to measure financial development [48]. Unlike developed countries with mature marketization, China is in a critical period of economic transformation, and the urban market mechanism is still imperfect. Local governments intervene relatively heavily in regional economic operations, usually pursuing goals such as economic growth and stable employment through fiscal spending, policy guidance, and resource allocation. Excessive intervention and regulation of economic activities by local governments lead to an obvious non-market-oriented feature in regional resource allocation; production factors cannot flow freely and allocate efficiently within and across cities in response to price changes. This may result in the excessive concentration of innovation resources in inefficient entities and fields, and also breed politically selective innovation behaviors and innovation outputs caused by power rent-seeking, which damages the sustained growth of regional innovation efficiency. By contrast, moderate institutional constraints can alleviate market segmentation and thereby improve resource allocation efficiency. Accordingly, this study also includes government intervention (intervention) as a control variable, measured by the ratio of local government fiscal expenditure to urban gross regional product.

3.2. Model Specification

3.2.1. Baseline Regression Model

To examine the direct impact of digital government on regional scientific and technological innovation capacity (H1), this paper constructs a two-way fixed-effects panel model for baseline estimation [37]. This model controls for time-invariant heterogeneous characteristics across cities and macro time-trend shocks, effectively mitigating omitted variable bias.

$$\mathrm{tech}_{it} = \beta_0 + \beta_1 \mathrm{dg}_{it} + \sum \beta_j \, \mathrm{CV}_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{1}$$

where tech_{it} denotes the regional scientific and technological innovation capacity of city i in year t; dg_{it} is the core explanatory variable (digital government); CV_{it} represents the set of control variables (human capital level, openness, industrial structure, financial development, and government intervention); μ_i is the city fixed effect; λ_t is the year fixed effect; ε_{it} is the random error ter; β_0 is the constant term; β_1 is the core coefficient; and β_j denotes the coefficients of the control variables.

3.2.2. Endogeneity Test Design

Although the two-way fixed-effects model has mitigated the endogeneity bias caused by omitted variables to some extent, a bidirectional causal relationship may still exist between digital government construction and regional technological innovation. On the one hand, digital government promotes regional innovation by optimizing government services and empowering factor allocation. On the other hand, cities with higher innovation vitality often possess more abundant fiscal resources and technical foundations, which in turn accelerate the process of digital government construction. This reverse causality may lead to biased estimates in the benchmark regression. Therefore, this paper further employs the Two-Stage Least Squares (2SLS) instrumental variable method for causal identification to ensure the reliability of the core conclusions.

Local governments in China exhibit a significant "peer effect" in the practice of administrative reform and digital governance [9]. This peer effect is specifically manifested

in the tendency of local governments to imitate and follow the behaviors of other local governments within the same province when making economically relevant decisions [6]. As a core institutional innovation for optimizing government services and empowering economic development, the construction process of digital government is also profoundly influenced by this peer effect. This characteristic provides a natural identification condition for constructing instrumental variables in this paper. Therefore, we select the mean value of digital government construction of other cities in the same province (excluding the city itself) in the same year as the instrumental variable for the core explanatory variable. Its rationality fully satisfies the two core assumptions of instrumental variables.

Regarding relevance, prefecture-level cities within the same province exhibit substantial homogeneity in institutional environments and policy alignment. Provincial-level mandates—such as digital government construction targets and integrated public service reforms—impose unified regulatory frameworks and developmental trajectories across jurisdictions. Concurrently, under the dual pressure of inter-city peer effects and tournament-style promotion competition, a positive association emerges: as the average level of digital governance in neighboring cities increases, the subject city faces intensified incentives to accelerate its own digital transformation. Consequently, the mean digital government index of other cities within the same province serves as a robust proxy, demonstrating a strong first-stage correlation with the focal city's digital governance level.

Regarding the exclusion restriction, the instrument derives from the digital governance levels of other cities within the province, isolating it from the endogenous decision-making processes of the focal city. The transmission channel operates strictly through the focal city's own digital government development; it does not exert a direct influence on local innovation inputs, such as corporate R&D expenditure, or outputs, such as patent generation. This mechanism satisfies the exogeneity requirement and upholds the exclusion restriction, effectively mitigating concerns regarding reverse causality and omitted variable bias. We further control for province-year fixed effects to absorb shared shocks. The corresponding first-stage and second-stage models are specified as follows:

$$\text{FirstStage}: \mathrm{dg}_{it} = \phi_0 + \phi_1 \overline{\mathrm{dg}}_{jt} + \sum \phi_j \, \mathrm{CV}_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{2}$$

$$\text{SecondStage}: \mathrm{tech}_{it} = \varphi_0 + \varphi_1 \hat{\mathrm{dg}}_{it} + \sum \varphi_j \, \mathrm{CV}_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{3}$$

where $\overline{\mathrm{dg}}_{jt}$ represents the instrumental variable (mean digital government level of other cities in province j in year t, excluding city i); $\hat{\mathrm{dg}}_{it}$ denotes the fitted values from the first-stage regression; ϕ_0 and φ_0 are the constant terms of the first-stage and second-stage models, respectively; ϕ_1 captures the effect of the instrumental variable on the focal city's digital government development; ϕ_j and φ_j denote the coefficients of control variables in the first and second stages, respectively; φ_1 is the core coefficient of interest representing the causal effect of digital government on regional scientific and technological innovation capacity; and all other variables dg_{it}, tech_{it}, CV_{it}, λ_t, ε_{it} are defined as in the baseline model.

3.2.3. Robustness Check Design

To ensure the reliability of the baseline conclusions, we design a series of robustness tests. First, considering the systematic shocks imposed by the COVID-19 pandemic, we will re-run the regression after excluding the samples from 2019 and 2020. Second, to avoid potential interference from outlier handling, we will replace the dependent variable with the logarithm of the number of unwinsorized granted patents. Third, given that digital government is a multidimensional construct, we will decompose it into three sub-dimensions—service supply capacity, service responsiveness, and service intelligence—and conduct separate and joint regressions to identify their differential impacts.

3.2.4. Mediating Effect Models

To test the mechanism paths through which digital government affects regional scientific and technological innovation capacity, models are constructed following the method proposed by relevant scholars [49].

Mediating Effect of Digital Industry Agglomeration

$$\mathrm{digagg}_{it} = \alpha_0 + \alpha_1 \mathrm{dg}_{it} + \sum \alpha_j \, \mathrm{CV}_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{4}$$

$$\mathrm{tech}_{it} = \gamma_0 + \gamma_1 \mathrm{dg}_{it} + \gamma_2 \mathrm{digagg}_{it} + \sum \gamma_j \, \mathrm{CV}_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{5}$$

where digagg_{it} is the level of digital industry agglomeration; α_0 and γ_0 are the constant terms of Equation (4) and Equation (5), respectively; α_1 is the coefficient of digital government's impact on digital industry agglomeration; α_j is the coefficient of control variables in Equation (4); γ_1 is the direct effect coefficient of digital government on technological innovation; γ_2 is the coefficient of digital industry agglomeration's impact on technological innovation; and γ_j is the coefficient of control variables in Equation (5). All other variables are defined as above.

Mediating Effect of Resource Misallocation

$$\mathrm{misall}_{it} = \delta_0 + \delta_1 \mathrm{dg}_{it} + \sum \delta_j \, \mathrm{CV}_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{6}$$

$$\mathrm{Tech}_{it} = \theta_0 + \theta_1 \mathrm{dg}_{it} + \theta_2 \mathrm{misall}_{it} + \sum \theta_j \, \mathrm{CV}_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{7}$$

where misall_{it} denotes the resource misallocation index (a larger value indicates more severe misallocation); δ_0 and θ_0 are the constant terms of Equations (6) and (7), respectively; δ_1 is the coefficient of digital government's impact on resource misallocation; δ_j is the coefficient of control variables in Equation (6); θ_1 is the direct effect coefficient of digital government on technological innovation; θ_2 is the coefficient of resource misallocation's impact on technological innovation; and θ_j is the coefficient of control variables in Equation (7). All other variables are defined as above.

3.2.5. Moderating Effect Model

To test the positive moderating effect of regional economic development level (H3), an interaction term is introduced to construct the model:

$$\mathrm{Tech}_{it} = \eta_0 + \eta_1 \mathrm{dg}_{it} + \eta_2 \mathrm{gdp}_{it} + \eta_3 (\mathrm{dg}_{it} \times \mathrm{gdp}_{it}) + \sum \eta_j \, \mathrm{CV}_{it} + \mu_i + \lambda_t + \varepsilon_{it} \tag{8}$$

where gdp_{it} represents the regional economic development level; $\mathrm{dg}_{it} \times \mathrm{gdp}_{it}$ is the interaction term between digital government and regional economic development level; η_0 is the constant term of the model; η_1 is the main effect coefficient of digital government; η_2 is the main effect coefficient of regional economic development level; η_3 is the moderating effect coefficient of the interaction term; and η_j is the coefficient of control variables. All other variables are defined as above.

4. Empirical Analysis

To mitigate the interference of extreme values on regression results, all continuous variables in this study are subjected to 1% two-sided winsorization, and all empirical analyses are conducted using the winsorized data.

4.1. Descriptive Statistics and Multicollinearity Test

4.1.1. Descriptive Statistics of Main Variables

Table 2 reports the descriptive statistics of the core variables, mediating variables, moderating variables, and control variables in this study. The full sample contains 1560 observations, and the distribution characteristics of each variable are as follows.

Table 2. Descriptive statistics of the main variables.

	Observations	Mean	P50	SD	Min	Max
Tech	1560	8.538	8.428	1.365	5.940	12.064
DG	1560	0.695	0.704	0.097	0.438	0.867
Digagg	1560	0.531	0.396	0.465	0.093	2.895
Misall	1560	0.317	0.283	0.193	0.026	0.928
GDP	1560	10.994	10.893	0.625	9.896	12.655
Human	1560	0.023	0.013	0.026	0.002	0.131
Opening	1560	0.002	0.001	0.002	0.000	0.011
Finance	1560	3.007	2.662	1.177	1.378	6.908
IS	1560	0.483	0.478	0.079	0.297	0.733
Intervention	1560	0.207	0.180	0.094	0.084	0.561

The dependent variable (regional scientific and technological innovation capacity) is measured by the logarithm of the number of patent grants, with a mean of 8.538 and a standard deviation of 1.365, indicating substantial disparities across cities. The core independent variable (digital government) has a mean of 0.695 and a standard deviation of 0.097, which is at a moderate level overall with gentle regional gaps. The moderating variable (regional economic development level) has a mean of 10.994 and a standard deviation of 0.622, showing remarkable differences in economic foundations among cities. For the mediating variables, the level of digital industry agglomeration has a mean of 0.531 and a standard deviation of 0.465, featuring uneven spatial distribution and prominent regional gaps; the resource misallocation index has a mean of 0.317, indicating a low overall distortion degree but sharp divergences in resource allocation efficiency across cities.

The statistical results of control variables (human capital level, opening-up level, financial development level, industrial structure, and government intervention) reveal that sample cities differ to varying degrees in talent endowment, openness, financial development, industrial structure, and government behavior. Among them, the overall levels of human capital and opening-up are relatively low, reflecting that most prefecture-level cities still have room to improve high-end factor agglomeration and external market access. The regional gap in financial development is relatively conspicuous, embodying the unbalanced spatial allocation of financial resources. The small standard deviation of industrial structure suggests moderate divergences in industrial structure among cities. The regional divergence in government intervention provides effective support for controlling heterogeneous disturbances in the model.

Overall, the distribution characteristics of the core variables, mediating variables, moderating variables, and control variables align with the development reality of prefecture-level cities in China. The variables present reasonable and sufficient internal disparities, which can provide reliable data support for subsequent benchmark regression, mechanism tests, moderating effect analysis, endogeneity tests, robustness checks, and heterogeneity analysis.

4.1.2. Multicollinearity Test

To avoid severe multicollinearity in the model and ensure the reliability of the regression results, this paper conducts the Variance Inflation Factor (VIF) test for all core

explanatory variables, mediating variables, moderating variables, and control variables. Given that this paper employs a two-way fixed-effects model to control for heterogeneous shocks at the individual and time levels, year dummy variables (with the base year as the reference group) are included in the test to fully reflect the multicollinearity level of the model variables. The detailed statistical results are reported in Table 3. The test results show that the VIF values of all variables are far below the empirical critical value of 10. The maximum VIF value is only 8.64, and the mean VIF value is merely 3.31, indicating that the model has no severe multicollinearity problem and satisfies the basic prerequisite assumptions for the baseline regression analysis.

Table 3. Multicollinearity test results.

Variable	VIF	1/VIF
DG	2.33	0.429709
Digagg	1.18	0.849683
Misall	1.37	0.728994
GDP	5.54	0.180551
Human	1.31	0.761558
Opening	1.05	0.948829
Finance	2.69	0.371567
IS	1.37	0.728914
Intervention	2.08	0.480392
Dummy variables in 2019	8.64	0.115747
Dummy variables in 2020	7.01	0.142593
Dummy variables in 2021	5.38	0.185981
Dummy variables in 2022	3.81	0.262527
Dummy variables in 2023	2.55	0.392291
Mean VIF	**3.31**	**-**

Notes: A VIF value below 10 indicates no severe multicollinearity. Year dummies take the base year of the sample as the reference group to control for time-fixed effects. Sample size number = 1560.

4.2. Baseline Regression Results

We perform the baseline regression using Equation (1), with the results reported in Table 4. Table 4 presents the baseline regression results of digital government's impact on regional scientific and technological innovation, with control variables and time fixed effects added sequentially. Model (1) is a fixed effects regression including only the core explanatory variable. The results show that the coefficient of digital government is significantly positive at the 1% level, preliminarily indicating that digital government construction markedly boosts regional scientific and technological innovation capacity. Model (2) incorporates a series of city-level control variables on the basis of Model (1). The coefficient of the core explanatory variable (digital government) remains significantly positive at the 1% level and declines only slightly, meaning that the empowering effect of digital government on innovation remains robust after controlling for other influencing factors. Model (3) further controls for time fixed effects to rule out interference from unobservable time-varying factors such as macroeconomic policy shocks and external environmental changes. The results show that the coefficient of digital government is 0.187 and significantly positive at the 5% statistical level. This finding empirically verifies the core theoretical hypothesis H1 of this paper: digital government construction effectively improves regional scientific and technological innovation capacity.

Table 4. The benchmark regression results.

Variable	Model (1) Core Variables Only	Model (2) Control Variables Added	Model (3) Two-Way Fixed Effects
DG	1.084 ***	0.835 ***	0.187 **
	(0.062)	(0.070)	(0.074)
Human	-	5.556 ***	1.439
		(1.126)	(0.939)
Opening	-	−13.543	0.398
		(8.707)	(7.006)
Finance	-	0.081 ***	−0.050 **
		(0.017)	(0.017)
IS	-	0.305 *	−0.665 ***
		(0.185)	(0.159)
Intervention	-	−0.075	0.140
		(0.207)	(0.202)
Constant	7.785 ***	7.481 ***	8.561 ***
	(0.043)	(0.099)	(0.091)
City FE	YES	YES	YES
Year FE	NO	NO	YES
Observations	1560	1560	1560
Within R^2	0.195	0.240	0.521

Notes: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Regarding the regression results of the control variables, the coefficient of financial development in the final baseline Model (3) is −0.050, significantly negative at the 5% statistical level, indicating that the expansion of financial development scale exerts a significant inhibitory effect on regional scientific and technological innovation capacity. This finding stands in stark contrast to the conventional theoretical expectation that "financial deepening fosters technological innovation." However, it accurately captures a stylized fact in transitional economies: the disconnect between the scale expansion of finance and its allocative efficiency. Extant literature argues that in bank-dominated financial systems, credit resources tend to flow preferentially toward large state-owned enterprises and traditional mature industries possessing implicit guarantees, while innovative small and medium-sized enterprises with high growth potential but insufficient collateral face severe financing constraints [49]. This capital misallocation, driven by ownership bias and collateral-based lending practices, reduces the marginal productivity of financial resources, thereby suppressing overall innovation efficiency. Furthermore, excessive financial expansion may lead to liquidity remaining trapped within the financial system for speculative arbitrage, or crowding out R&D resources from the real economy through credit rent-seeking behaviors—manifesting as the "crowding-out effect" of "excessive financialization" on the real economy [50]. Therefore, the negative coefficient observed in this study should be interpreted as follows: financial deepening that relies solely on scale expansion, without a simultaneous improvement in credit allocation mechanisms based on innovation risk assessment and total factor productivity, fails to translate into a driver of regional innovation.

The coefficient of industrial structure is significantly negative at the 1% level, suggesting that industrial structure upgrading inhibits regional scientific and technological innovation after controlling for individual and time fixed effects. This does not imply that the development of the service sector inherently damages innovation. Rather, it serves as a critical caveat: using the "rising share of the tertiary sector" as a proxy for structural upgrading essentially captures a change in the compositionof industries, not the quality

of upgrading. When this shift is primarily driven by the expansion of low-value-added traditional services, it is not accompanied by enhanced externalities from advanced producer services on manufacturing R&D [29], but merely manifests as a "deindustrialization appearance" in the economic structure. Given that our dependent variable (patent grants) is more closely tied to the manufacturing/industrial R&D system, this change represents an "illusion of upgrading" rather than an enhancement of "upgrading capacity"—hence its observed negative conditional relationship. This finding resonates with concerns regarding the adverse context of "premature deindustrialization" for the innovation foundation [47]. In other words, the negative coefficient should be interpreted as follows: merely increasing the share of the tertiary sector is insufficient; genuine innovation-conducive upgrading requires the synergistic advancement of high-end producer services and the manufacturing technology system.

The coefficients of human capital, openness, and government intervention are all statistically insignificant, indicating that their effects on innovation are no longer significant after controlling for individual and time fixed effects. The insignificance of human capital may stem from the mismatched human capital structure and insufficient supply of high-end innovative talents in some cities, where the improvement of overall labor quality fails to translate effectively into innovation momentum. The insignificance of openness reflects the diminishing marginal effect of the traditional foreign capital-driven innovation model under the new development pattern with the domestic cycle as the mainstay. The insignificance of government intervention implies that direct government fiscal investment has a limited direct pulling effect on innovation, requiring further optimization of the methods and efficiency of government intervention, shifting from direct subsidies to institutional environment construction and innovation ecosystem cultivation.

Overall, the baseline regression results fully verify H1: digital government construction significantly and positively improves urban scientific and technological innovation capacity.

4.3. Endogeneity Test

Consistent with the identification strategy outlined in Section 3.2.2, we address potential endogeneity concerns—specifically reverse causality—by implementing a Two-Stage Least Squares (2SLS) instrumental variable estimation. Table 5 reports the regression results of the 2SLS instrumental variable method. The first-stage regression shows that the coefficient of the instrumental variable (provincial average) on the core explanatory variable (digital government) is 0.982, which is highly significant at the 1% statistical level. The first-stage F-statistic is 394.86, far exceeding the critical value of 10 for instrumental variable validity, strongly rejecting the hypothesis of weak instrumental variables. This indicates that the instrumental variable is strongly correlated with the endogenous explanatory variable, and the instrumental variable setting is valid. The second-stage regression results show that after controlling for endogeneity bias, the coefficient of the core explanatory variable (digital government) is 0.508, which remains significantly positive at the 1% level. Comparing this with the benchmark regression results, we find that after addressing reverse causality and endogeneity issues, the sign of the coefficient of digital government on technological innovation remains consistent with the benchmark regression, and the significance level is stable. This indicates that even after considering endogeneity disturbances such as reverse causality and measurement errors, digital government construction can still significantly enhance urban technological innovation capability. The core conclusion of this paper possesses solid causal identification power.

Table 5. Endogeneity test results.

Variable	First-Stage Regression: The Dependent Variable Is Digital Government.	Second-Stage Regression: The Dependent Variable Is Technological Innovation
Provincial average	0.982 *** (0.049)	-
DG	-	0.508 *** (0.152)
Human	0.300 (0.312)	1.326 (0.948)
Opening	0.800 (2.330)	0.079 (7.060)
Finance	0.010 ** (0.006)	−0.057 *** (0.018)
IS	−0.063 (0.053)	−0.665 *** (0.160)
Intervention	0.008 (0.067)	0.181 (0.204)
City FE	YES	YES
Year FE	YES	YES
Observation	1560	1560
Within R^2	0.680	-
First-stage F-statistic	394.86	-
Second-stage Wald χ^2	-	5.98×10^6

Notes: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4.4. Robustness Tests

To further verify the reliability of the baseline regression conclusions, this paper conducts a series of robustness tests from three dimensions: excluding special shocks, replacing core variables, and dimensional decomposition, with the results presented in Table 6.

Table 6. Robustness test results.

Variable	Model (1) Excluding Pandemic Shock	Model (2) Replacing Dependent Variable	Model (3) Service Supply Capacity	Model (4) Service Responsiveness	Model (5) Service Intelligence	Model (6) Three-Dimensional Joint Regression
DG	0.299 *** (0.109)	0.236 ** (0.083)	-	-	-	-
Supply	-	-	0.382 ** (0.153)	-	-	0.379 ** (0.153)
Responsiveness	-	-	-	0.147 (0.101)	-	0.142 (0.101)
Intelligence	-	-	-	-	0.200 (0.176)	0.150 (0.177)
Human	1.511 (1.150)	1.190 (1.057)	1.497 (0.939)	1.495 (0.941)	1.474 (0.941)	1.467 (0.939)
Opening	9.150 (8.771)	3.020 (7.883)	0.687 (7.006)	0.193 (7.023)	0.722 (7.021)	0.411 (7.010)
Finance	−0.063 *** (0.024)	−0.038 * (0.019)	−0.051 *** (0.017)	−0.045 ** (0.017)	−0.049 *** (0.017)	−0.051 *** (0.017)

Table 6. *Cont.*

Variable	Model (1) Excluding Pandemic Shock	Model (2) Replacing Dependent Variable	Model (3) Service Supply Capacity	Model (4) Service Responsiveness	Model (5) Service Intelligence	Model (6) Three-Dimensional Joint Regression
IS	−0.589 ***	−0.431 **	−0.680 ***	−0.664 ***	−0.659 ***	−0.674 ***
	(0.195)	(0.179)	(0.159)	(0.159)	(0.159)	(0.159)
Intervention	0.163	−0.389 *	0.183	0.107	0.113	0.171
	(0.292)	(0.227)	(0.203)	(0.202)	(0.202)	(0.204)
City FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Constant	8.481 ***	8.081 ***	8.564 ***	8.631 ***	8.650 ***	8.521 ***
	(0.123)	(0.102)	(0.090)	(0.084)	(0.082)	(0.094)
Observation	1028	1560	1560	1560	1560	1560
Within R^2	0.529	0.795	0.521	0.519	0.519	0.522

Notes: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4.4.1. Excluding the Impact of the COVID-19 Pandemic

The COVID-19 pandemic in 2019–2020 exerted systematic shocks on urban economic operations and innovation activities in China, which may interfere with the identification of the core conclusions. To address this issue, this paper re-runs the regression after excluding the samples of 2019 and 2020. The results show that the coefficient of digital government is significantly positive at the 1% statistical level, with the sign and direction consistent with the baseline regression conclusions, indicating that the pandemic shock does not alter the promoting effect of digital government on scientific and technological innovation, and the core conclusions are robust and reliable.

4.4.2. Replacing the Dependent Variable

To avoid the interference of outlier handling methods on the baseline regression results, this paper re-runs the regression using the logarithm of the number of unwinsorized granted patents (lnGranted) as the alternative indicator for scientific and technological innovation. The results show that the coefficient of the core explanatory variable (digital government) is 0.236, significantly positive at the 5% statistical level, which is consistent with the baseline regression results. This demonstrates that the promoting effect of digital government on regional scientific and technological innovation is not affected by outlier handling methods, and the core conclusions of this paper have high robustness and reliability.

4.4.3. Dimensional Test of the Core Explanatory Variable

The results of the single-dimensional regression show that only the coefficient of service supply capacity is significantly positive at the 5% level, while the coefficients of service responsiveness and service intelligence fail to pass the significance test. This indicates that the empowering effect of digital government on innovation is mainly realized through the core path of optimizing government service supply and improving government service efficiency, and the innovation-empowering roles of responsiveness and intelligence have not been fully exerted.

The results of the three-dimensional joint regression show that the coefficient of service supply capacity remains significantly positive at the 5% level, while the coefficients of the other two dimensions are insignificant, which is consistent with the single-dimensional regression conclusions and further verifies the robustness of the dimensional results.

In summary, the series of robustness tests confirms that the significant promoting effect of digital government on regional scientific, and technological innovation is robust and reliable, free from the interference of special shocks, variable measurement methods and other factors. The core conclusions of this paper have high credibility and persuasiveness.

4.5. Mediation Effect Tests

Adhering to the prevailing mainstream paradigm for testing mediation effects, this study abandons the traditional stepwise regression approach. Instead, we employ a two-way fixed-effects model combined with a Bootstrap sampling method (5000 repetitions) to estimate the 95% percentile confidence intervals for the indirect effects. The validity of the mediation channels is determined by whether the confidence intervals exclude zero, thereby effectively circumventing the limitations of stepwise regression—namely, inflated Type I error rates and insufficient statistical power. The results of the mediation analysis are presented in Table 7.

Table 7. Bootstrap test results of indirect mediation effects.

Mediator Variable	Indirect Effect Coefficient	Bootstrap Std. Error	95% Percentile Confidence Interval	Mediation Conclusion
Digital Industrial Agglomeration	0.000467	0.000332	[−0.000137, 0.001241]	95% CI marginally includes zero, partial mediation valid at 10%
Resource Misallocation	0.005054	0.003868	[−0.001155, 0.014304]	Left bound of 95% CI close to zero, partial mediation valid at 10%

4.5.1. Digital Industry Agglomeration

Grounded in the "Reintegration" logic of Digital-Era Governance (DEG) theory, digital government promotes digital industrial agglomeration by dismantling bureaucratic barriers, thereby driving regional scientific and technological innovation. The Bootstrap test results (see Table 7) indicate that the indirect effect of digital industry agglomeration is 0.000467, with a 95% confidence interval of [−0.000137, 0.001241]. While the interval marginally includes zero, the partial mediation effect is supported at the 10% significance level (Hypothesis H2a is confirmed).

Although the statistical significance of this pathway is relatively weak (marginally significant), its economic substance is noteworthy. This result suggests that digital government primarily fosters industrial agglomeration through "institutional provision" rather than "direct intervention." Specifically, by establishing data-sharing platforms and streamlining administrative approval processes, digital government lowers the entry barriers and transaction costs for digital firms, thereby catalyzing innovation clusters. However, the modest effect size (the indirect effect accounts for only approximately 0.26% of the total effect) reveals a characteristic "running-in period" in current digital government initiatives: a protracted transmission chain persists between the perfection of institutional infrastructure and actual industrial agglomeration, which is highly susceptible to dilution by the region's pre-existing industrial ecosystem.

4.5.2. Mediating Effect of Resource Misallocation

Grounded in the "Needs-based Holism" logic of DEG theory, digital government alleviates resource misallocation through the precise matching of factor demands, thereby

unlocking innovation potential. The Bootstrap test results (see Table 7) indicate that the indirect effect of resource misallocation alleviation is 0.005054, with a 95% confidence interval of [−0.001155, 0.014304]. Although the left bound of the interval is close to zero, the existence of a partial mediation effect is supported at the 10% significance level (Hypothesis H2b is confirmed).

The effect size of this pathway is significantly larger than that of the industrial agglomeration pathway, indicating that "removing impediments" constitutes the core mechanism through which digital government currently empowers innovation. Specifically, by constructing a unified market access and regulatory system, digital government dismantles local protectionism and hidden barriers to factor mobility, enabling capital, talent, and other critical innovation factors to concentrate in high-efficiency sectors. The marginal returns of this "reform dividend" far exceed those of mere "technological input," explaining why the innovation-enhancing effect of digital government tends to be more pronounced in regions with rigid factor markets.

4.6. Moderating Effect Analysis

The innovation-enabling effect of digital government does not exist in isolation; rather, it is nested within the macro-structure of regional socio-economic systems. Grounded in the "Needs-based Holism" logic of Digital-Era Governance (DEG) theory, digital government serves as a meta-technology for reconstructing governance systems, and its efficacy is highly contingent on the supporting resource endowments provided by regional economic development levels. Regional economic development represents not only infrastructure and market scale but also constitutes the structural scaffolding that enables the absorption and transformation of digital governance technologies.

The regression results in Table 8 corroborate this dependency. The coefficient of the interaction term between digital government (DG) and regional economic development level (GDP) is 0.163, which is statistically significant at the 5% level. This result confirms that the impact of digital government on innovation is not fixed but exhibits a significant positive moderating effect as economic development levels increase.

To intuitively unpack the substance of this interaction effect, this study plots the marginal effect graph of digital government under varying levels of economic development (see Figure 2). The solid line represents the trajectory of digital government's direct impact on innovation under different economic conditions. At lower levels of economic development, the marginal effect of digital government approaches zero or shows a weak negative trend, with the corresponding confidence intervals (vertical error bars) straddling zero. This suggests that the innovation dividends of digital government construction have yet to materialize at this stage and may even lead to a crowding-out effect due to insufficient supporting infrastructure. Conversely, as economic development enters the upper-middle stage, the marginal effect curve rises significantly and turns positive, with the confidence interval narrowing progressively. This indicates that the innovation-enabling mechanism of digital government—operating through optimizing resource allocation and reducing transaction costs—begins to function effectively, yielding positive and robust marginal returns.

Further validation of the interaction effect is conducted through Simple Slope Analysis (see Figure 3). The sample is divided into high- and low-economic development groups based on one standard deviation above and below the mean. The results show that the slope of the fitted line for the high-economic group is significantly steeper than that for the low-economic group. This visually demonstrates that in regions with higher economic development, the driving effect of digital government construction on scientific and technological innovation is more pronounced, whereas in less-developed regions, the

marginal contribution of digital government is relatively flat. This finding aligns perfectly with the trends observed in the marginal effects plot, collectively establishing a chained explanatory logic: "Economic Development Level-Maturity of Factor Supply-Efficiency of Digital Government Efficacy Transformation."

Table 8. Moderating effect test results.

Variable	Regional Scientific and Technological Innovation
DG	−1.600 *
	(0.901)
GDP × DG	0.163 **
	(0.082)
Human	1.300
	(0.941)
Opening	−0.659
	(7.008)
Finance	−0.074 ***
	(0.019)
IS	−0.738 ***
	(0.160)
Intervention	−0.165
	(0.217)
City FE	YES
Year FE	YES
Constant	12.338 ***
	(1.027)
Observation	1560
Within R^2	0.526

Notes: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

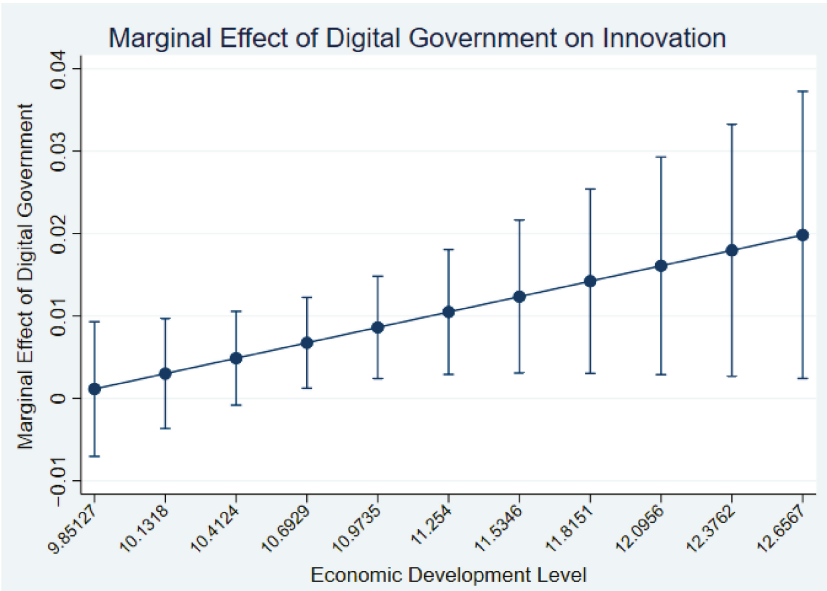


Figure 2. Marginal effect of digital government on innovation.

These results reveal the contextual dependency of digital government's innovation effect. Economic development enhances the regional system's absorptive capacity and transformation efficiency in digital governance technologies by providing mature factor markets and institutional environments. This co-evolutionary relationship indicates that the success of digital government construction is not merely a matter of technological

input but is deeply embedded in a region's specific development stage. Mere technological accumulation cannot automatically translate into innovation momentum; only when the economic system possesses sufficient carrying capacity and maturity can the "institutional dividends" of digital government be effectively activated and converted into tangible innovation outputs.

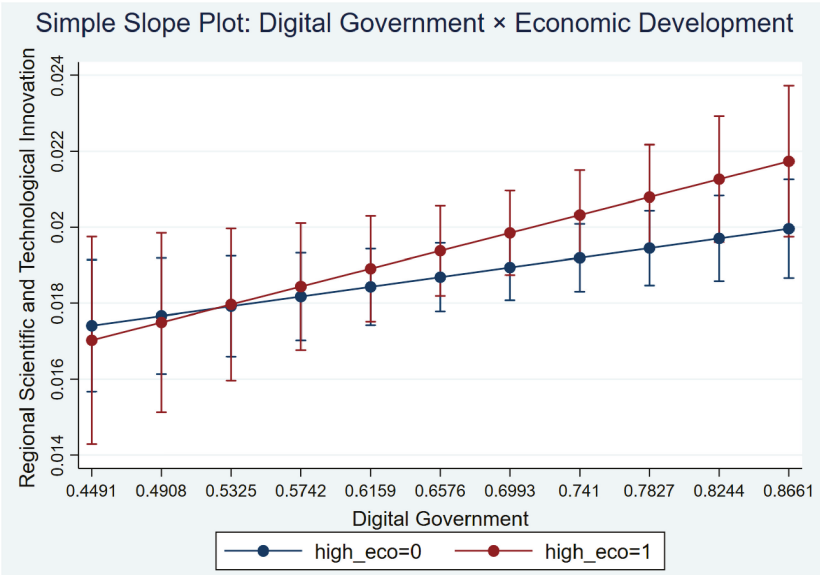


Figure 3. Simple slope plot: digital government × economic development.

This study, based on a linear interaction model, identifies the positive moderating role of economic development on the innovation effect of digital government. Although the data reveal a clear trend of increasing marginal effects, it is important to note cautiously that the current model specification does not include a formal nonlinear threshold test. Therefore, the focus of this analysis is on confirming the direction and statistical significance of the moderation, rather than pinpointing specific critical thresholds. This finding validates Hypothesis H3 and provides empirical evidence for understanding the co-evolution within complex socio-technical systems.

4.7. Heterogeneity Tests

Digital-Era Governance (DEG) theory posits that digital government reconstructs the governance interface through "Reintegration," "Needs-based Holism," and "Digitization" to reduce administrative friction and enhance the allocation efficiency of innovation factors. However, the actual efficacy of this governance intervention is not homogeneous but exhibits significant contextual dependency. Drawing on the canonical structural features of regional innovation systems (RIS), this study examines the implementation variance of DEG mechanisms across different city subsamples along three dimensions: industrial base, network density, and scale thresholds. It is crucial to emphasize that this heterogeneity analysis aims to uncover differences in marginal returns of the same governance reform across varying initial conditions, rather than testing the causal moderation of a variable on the main effect. This approach helps circumvent estimation biases arising from omitted variables or sample selection.

4.7.1. Resource-Based vs. Non-Resource-Based Cities

The heterogeneity test results comparing resource-based and non-resource-based cities are presented in Table 9. The innovation effect of digital government is significantly positive in non-resource-based cities (0.278, $p < 0.01$), whereas it remains statistically insignificant in

resource-based cities. This divergence stems from the constraints imposed by factor market distortions on the core mechanisms of DEG.

Table 9. Heterogeneity test results: resource-based vs. non-resource-based.

Variable	Resource-Based Cities	Non-Resource-Based Cities
DG	0.080 (0.111)	0.278 *** (0.099)
Human	2.919 (2.217)	1.100 (1.033)
Opening	11.255 (13.130)	−4.042 (8.295)
Finance	−0.467 (0.030)	−0.053 * (0.021)
IS	−0.696 *** (0.236)	−0.638 *** (0.222)
Intervention	−0.034 (0.290)	0.309 (0.300)
City FE	YES	YES
Year FE	YES	YES
Constant	7.912 *** (0.126)	8.977 *** (0.133)
Observation	632	928
Within R^2	0.537	0.513

Notes: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The diversified industrial structure of non-resource-based cities provides a compatible environment for the "Reintegration" mechanism, where the dismantling of data silos can be rapidly translated into optimized resource allocation [10]. In contrast, resource-based cities are often constrained by the institutional rigidity of traditional industrial systems. Under the "resource-dependent" development paradigm, local government fiscal revenue, credit allocation, and industrial policies are highly concentrated on a few dominant industries, forming a stable symbiosis of vested interests. This structure locks critical innovation factors—such as capital and talent—into established industrial trajectories, hindering their flow toward emerging industries. In this context, the technological embedding of digital government is often absorbed by existing administrative inertia. Its functionality remains largely confined to the digital replication and efficiency improvements of traditional approval processes, failing to achieve the deeper objectives of industrial chain restructuring and cross-sector innovation synergy. Consequently, this blocks the substantive translation of governance logic into innovation outputs.

4.7.2. Central vs. Non-Central Cities

The regression results distinguishing central cities from non-central cities are presented in Table 10. The innovation effect of digital government is statistically significant only in non-central cities (coefficient = 0.177, $p < 0.05$), while it remains insignificant in central cities. This phenomenon reflects the differential marginal returns of "Digitization" under varying network densities. As mature regional innovation hubs, central cities already possess highly developed agglomeration effects in physical space and sophisticated intermediary organizations, where the transaction costs of market self-organization are already low. In this context, digital government primarily manifests as a marginal optimization of existing processes, yielding relatively limited innovation returns. In contrast, non-central cities commonly face the predicament of sparse innovation factors and a scarcity of institutional connections. The unified platforms constructed by digital government transcend geographical and administrative boundaries, connecting discrete innovation actors at a low cost,

thereby substantially compensating for the sparsity of offline networks. Consequently, "Digitization" yields significantly higher barriers to innovation. For marginal returns in these systems, digital government acts as a crucial systemic connector and market augmenter. It compensates for institutional thinness by providing digital platforms that lower the exceptionally high transaction costs of partner search and information access, thereby triggering a more pronounced innovation response in regions.

Table 10. Heterogeneity test results: central cities vs. non-central cities.

Variable	Central Cities	Non-Central Cities
DG	0.206 (0.162)	0.177 ** (0.080)
Human	0.108 (1.292)	1.547 (1.151)
Opening	24.726 (15.718)	−2.397 (7.599)
Finance	−0.057 * (0.031)	−0.050 ** (0.023)
IS	−1.256 *** (0.408)	−0.595 *** (0.170)
Intervention	1.582 ** (0.783)	0.039 (0.215)
City FE	YES	YES
Year FE	YES	YES
Constant	9.433 *** (0.147)	8.290 *** (0.094)
Observation	186	1374
Within R^2	0.649	0.480

Notes: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4.7.3. Large/Medium-Sized vs. Small Cities

The regression results by city scale (Table 11) show that the coefficient on digital government is substantially larger and highly significant in large and medium-sized cities (0.329, $p < 0.01$), whereas it is statistically insignificant in small cities. This pattern reveals a threshold effect of "Needs-based Holism" tied to demand scale and feedback density. In large and medium-sized cities, the sizable population base and richer application scenarios provide the necessary conditions for a needs-oriented governance logic to operate: a broad user base spreads platform operation and maintenance costs, while heterogeneous demand scenarios continuously force the iterative refinement of government service processes.

By contrast, small cities—constrained by narrower application scenarios and tighter fiscal space—tend to see their digital government platforms degenerate into single-purpose online approval tools. Lacking sufficient user traffic and the data feedback loops needed for learning, digital government struggles to accurately identify and match the heterogeneous needs of innovators. Consequently, the "demand orientation" becomes largely pro forma, and the deeper productivity potential of Digitization stays inert because the real economy lacks the carrying capacity/absorption base required to put it to work.

Table 11. Heterogeneity test results: large/medium-sized cities vs. small cities.

Variable	Large/Medium-Sized Cities	Small Cities
DG	0.329 *** (0.099)	0.103 (0.010)
Human	0.474 (0.893)	3.789 * (1.931)
Opening	−2.431 (9.041)	5.423 (9.798)
Finance	−0.042 * (0.025)	−0.067 *** (0.023)
IS	−0.622 ** (0.025)	−0.545 *** (0.222)
Intervention	1.304 (0.396)	−0.172 (0.248)
City FE	YES	YES
Year FE	YES	YES
Constant	9.433 *** (0.147)	7.970 *** (0.029)
Observation	574	986
Within R^2	0.649	0.480

Notes: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5. Conclusions and Policy Implications

5.1. Conclusions

Based on panel data from 280 Chinese cities spanning 2018 to 2023, this study moves beyond the traditional "technological instrumentalism" perspective. By situating digital government (DG) within the theoretical framework of Digital-Era Governance (DEG), we empirically examine its reshaping effect on urban scientific and technological innovation capacity. The findings not only confirm the reconfiguring role of digital government as a novel governance paradigm within the innovation system but also unveil the complexity and boundary conditions underlying this effect through mechanism identification and heterogeneity analysis. The principal conclusions are as follows:

First, digital government constitutes an institutional force for reconstructing regional innovation ecosystems. Empirical tests demonstrate that digital government significantly enhances regional scientific and technological innovation capacity at the aggregate level. This conclusion remains robust after addressing endogeneity concerns and passing multiple robustness checks. This finding confirms that DG is not merely an electronic extension of administrative processes; rather, it represents a critical institutional force that reconfigures regional relations of production through "Reintegration" and "Digitization." It signals a paradigmatic shift in governance from the traditional bureaucracy to the digital era.

Second, digital government empowers innovation through dual pathways: "Industrial Agglomeration" and "Resource Misallocation Alleviation." Mechanism analysis reveals that digital government drives innovation via dual mediation channels. Specifically, digital government fosters scale effects by attracting digital industrial agglomeration while simultaneously correcting factor allocation distortions by reducing institutional friction. Decomposition of the mediation effects shows that alleviating resource misallocation (aligned with the logic of "Needs-based Holism") contributes relatively more to the marginal effect. This implies that, at the current stage of development, digital government unleashes innovation potential more effectively by dismantling administrative barriers and enhancing the market-based allocation efficiency of factors, yielding stronger marginal returns than simply building industrial clusters. Nevertheless, both pathways serve as vital channels for the effectiveness of digital governance.

Third, regional economic development level acts as a crucial moderator of the innovation effect of digital government. Moderation effect tests indicate that regional economic development significantly moderates the innovation effect of digital government. This implies that digital government does not exert uniform force under all conditions; the release of its institutional dividends is highly contingent upon the regional economic foundation. The marginal promotion effect of digital governance on innovation is stronger in regions with a more robust economic base, whereas it weakens in less-developed areas. This finding establishes the economic foundation as a critical "supporting substrate" for the efficacy of digital governance.

Fourth, the innovation effect of digital government exhibits significant structural asymmetries. Heterogeneity analysis reveals differentiated manifestations of the innovation dividend. The innovation dividend of digital government is significantly positive in non-resource-based, non-central, and large-to-medium-sized cities. Conversely, in resource-based cities, central cities, and small cities, this effect is either systematically absorbed or rendered statistically insignificant. This suggests that, beyond the continuous moderating variable of economic foundation, structural conditions such as the pre-existing regional industrial ecosystem, network density, and scale thresholds also constitute significant determinants of digital government efficacy.

5.2. Policy Implications

First, prioritize consolidating "service supply capacity" and steer the transformation of digital government from "technology-centeredness" to "institution-centeredness." Empirical results based on dimensional decomposition (Table 6) indicate that the innovation-enabling effect of digital government stems primarily from service supply capacity, whereas service responsiveness and intelligence have yet to demonstrate significant impacts. This evidence suggests that relying solely on cutting-edge technology accumulation or interface optimization yields low input–output efficiency. Consequently, policy focus must pivot from the broad expansion of infrastructure back to core institutional restructuring, strictly adhering to the "Reintegration" logic of Digital-Era Governance (DEG) theory. Specifically, local governments should prioritize the construction of cross-departmental data-sharing platforms and the adoption of unified government service standards, while dismantling administrative barriers in key areas such as market regulation, taxation, and social security. This reduces the compliance and transaction costs for innovators by optimizing institutional interfaces. Simultaneously, evaluation systems must shift from process-oriented metrics—such as "online processing rates" or "platform coverage"—to outcome-based indicators, such as "improvement in factor allocation efficiency" and "enhancement of innovation output." Policies must ensure that digital governance translates concretely into public goods accessible to innovators, rather than remaining a mere digital replication of administrative workflows.

Second, implement a "dual-driver" strategy to alleviate resource misallocation. Since alleviating resource misallocation yields a higher marginal contribution than fostering digital industrial agglomeration, the allocation of policy resources should follow a "correction-first" logic. Regarding the "resource misallocation alleviation" pathway, digital government should be leveraged to unblock institutional bottlenecks hindering factor mobility, eliminate "ownership discrimination" and collateral dependency in credit allocation, and curb the inefficient circulation of financial resources within the system. Concurrently, digital technologies should be used to monitor and rationalize pricing mechanisms for key factors such as land and talent, eliminating distorted price signals to ensure that innovation factors flow dynamically from low- to high-efficiency sectors in line with efficiency principles. Regarding the "digital industrial agglomeration" pathway, avoid the trap of physical space

accumulation and "policy-induced agglomeration." Instead, leverage digital platforms to strengthen knowledge spillovers and digital collaboration across industrial chains within agglomeration zones. Construct low-cost information-sharing and transaction networks among industry, academia, and research institutions to prevent the hollowing-out phenomenon of "mere enterprise clustering without innovation linkage." Only by organically integrating "institutional correction in factor markets" with "networked reconstruction of innovation ecosystems" can the strongest driving force of digital governance on regional scientific and technological innovation be realized.

Third, implement a gradient advancement strategy calibrated to the moderating effect of economic foundations. Given the significant positive moderating role of regional economic development levels on the innovation effect of digital government, a one-size-fits-all national standard for digital construction should be avoided. In regions with a robust economic foundation, encourage the use of digital governance to explore frontier technological breakthroughs and cultivate future industries, leveraging high-factor mobility to achieve cross-domain innovation synergy. In regions with a relatively weaker economic base, priorities should focus on consolidating digital infrastructure and governance capacity, advancing large-scale governance projects prudently to prevent wasteful fiscal expenditure and distortions in innovation incentives caused by blindly premature construction.

Fourth, adopt classified governance strategies tailored to heterogeneous characteristics. Addressing the structural asymmetries in the innovation effects of digital government requires precise, categorized interventions. For resource-based cities, efforts must focus on dismantling the institutional rigidity induced by the existing industrial structure. Utilize digital technologies to build industrial transition support systems, prioritizing the removal of hidden barriers impeding factor flow toward emerging industries, thereby preventing digital government initiatives from being absorbed by traditional administrative inertia. For central cities, where the marginal returns of basic government service efficiency are diminishing, the governance focus should shift from mere process optimization to supplying institutional interfaces for complex innovation systems, emphasizing coordination for cross-regional innovation consortia and the sharing of major scientific infrastructure. For small cities, policymakers must acknowledge the constraints imposed by insufficient market scale on platform efficacy. The policy focus should shift from physical platform construction to building cross-regional digital innovation collaboration hubs, enabling low-cost access to external high-end resources. Ultimately, only by organically combining "institutional correction in factor markets" with "networked reconstruction of innovation ecosystems" can the strongest driving force of digital governance on regional scientific and technological innovation be achieved.

6. Contributions and Discussion

6.1. Contributions

First, this study extends the applicability of DEG theory within regional innovation research. Extant literature on digital government and innovation largely adheres to traditional public administration perspectives, reducing digital government to a technological tool variable or confining discussions to the singular dimension of governance efficiency [11]. This research breaks from these limitations by positioning the triple logic of DEG—Reintegration, Needs-based Holism, and Digitization—as the core explanatory mechanisms. This shift in perspective not only responds to Dunleavy et al.'s (2006) Digital-Era Governance (DEG) framework, which shifts the lens from NPM-style instrumental rationality to reintegration, needs-based holism, and digitization [20], but also furnishes a novel institutional explanatory variable for understanding regional disparities in innovation capacity. By moving beyond the conventional "input-output" analytical framework [1],

this study significantly broadens the empirical relevance of DEG theory in the domain of regional economic development.

Second, it unpacks the "black box" of digital government's enabling effect on innovation, clarifying the parallel pathways through which institutional advantages translate into innovation performance. While Lu et al. and Kalathil et al. confirm the positive impact of digital industrial agglomeration on innovation [28,29], and Hsieh & Klenow lay the theoretical groundwork for resource misallocation [32], existing studies fail to identify the specific institutional mechanisms through which digital government drives these processes. Departing from the vague narratives of previous research that vaguely attributed transmission mechanisms to "business environment improvement", this study simultaneously identifies and compares the dual mediation pathways of "digital industrial agglomeration" and "alleviation of resource misallocation." This finding not only extends Porter's (1990) theory of agglomeration externalities to the context of institutionally induced mechanisms in the digital era [2] but also addresses Brandt et al.'s concerns regarding factor market distortions in transitional economies [34]. It clarifies the micro-mechanisms through which digital government unlocks innovation potential by restructuring factor allocation rules rather than merely supplying technology.

Third, it identifies the boundary conditions for the efficacy of digital governance, reconciling academic debates concerning the drivers of regional innovation divergence. Addressing the "east-high-west-low" divergence trend highlighted in the Introduction, prior research predominantly attributes this phenomenon to infrastructure gaps [37] or differences in macro-institutional environments often overlooking the synergistic alignment between governance paradigms and economic development levels. By introducing absorptive capacity theory [40], this study empirically verifies that regional economic development positively moderates the innovation-enabling effect of DEG. This insight not only engages in dialog with North's (1990) thesis that "institutional efficacy is contingent upon economic fundamentals" [4] but also explains why the mere transplantation of technology fails to replicate the "innovation miracles" of the eastern region in the less-developed central and western areas. By elucidating the "economic foundation—governance paradigm—innovation output" logic chain, this research remedies the deficiency in traditional RIS literature regarding the neglect of digital dynamic capabilities.

Fourth, it reveals the urban-level heterogeneity of digital government's enabling effect on regional scientific and technological innovation. Countering the limitations of existing studies that often report national average effects, this research systematically characterizes the distribution of digital innovation dividends across different city types through three independent heterogeneity tests. The results indicate that the enabling effect is significantly positive only in non-resource-based, non-central, and large-to-medium-sized cities, whereas it is either systematically absorbed or rendered statistically insignificant in resource-based, central, and small cities. This discovery dispels the "mean illusion" of the undifferentiated diffusion of digital governance effects and responds to Cooke et al.'s discussions on regional heterogeneity and innovation-carrying capacity. It reveals the heterogeneous constraints imposed by pre-existing regional industrial ecosystems, network densities, and scale disparities on the efficacy of digital governance [23,24].

6.2. Limitations and Future Research Directions

First, limitations in measuring the dependent variable limit a comprehensive characterization of scientific and technological innovation capacity. This study employs the number of urban patent grants as a proxy for regional scientific and technological innovation capacity. This approach is deeply rooted in the innovation economics literature; patent data are regarded as a reliable indicator of technological output due to their high standardization,

broad coverage, and screening via substantive examination [51], and remain among the most frequently used measures in regional innovation research [52]. However, it must be acknowledged that patent grants essentially reflect patentable technological outputs and do not fully capture all dimensions of innovative activity. On the one hand, a substantial volume of innovation outcomes exists in non-patented forms—for instance, proprietary products or trade secrets that do not enter the patent system [53]. The omission of such "non-patentable innovation" leads to a systematic underestimation of innovative activity by patent counts alone [54]. On the other hand, the granting of a patent does not equate to its commercialization or generation of economic value [55]. Relying solely on patent counts makes it difficult to capture the critical translation phase from "technological output" to "economic performance." Future research could retain patent counts as a core metric while introducing supplementary multidimensional indicators: for example, incorporating economic performance metrics such as new product sales revenue or total factor productivity, or constructing a comprehensive evaluation system encompassing R&D input, patent output, and economic transformation [56]. Additionally, researchers could differentiate patent types (e.g., invention patents versus utility models) or assign weights based on citation data to provide a more nuanced characterization of regional scientific and technological innovation capacity.

Second, the mechanistic exploration of the negative effects of financial development and industrial structure remains insufficient. This study finds that the scale of financial development and the upgrading of the industrial structure (the share of the tertiary industry) exert significant negative impacts on regional scientific and technological innovation. This result diverges from the empirical expectations of "financial deepening promoting growth" and "structural transformation driving upgrading," revealing the prevalent disconnect between "scale and efficiency" and the mismatch between "structure and quality" in transitional economies. Drawing on the extant literature, we preliminarily attribute this to "ownership bias and collateral dependency" under a bank-dominated system and to the risk of "premature deindustrialization" during service-sector expansion. However, constrained by the granularity of macro-level city panel data, this study fails to unpack the internal "black box" of these two negative pathways. Regarding the financial dimension: does the expansion of credit scale inhibit innovation primarily through credit misallocation driven by ownership discrimination, maturity mismatch caused by short-term arbitrage, or capital idling resulting from decoupling from the real economy? Regarding the industrial dimension: does the negative effect of an increasing tertiary sector share stem from the direct crowding-out of manufacturing innovation resources by low-value-added services, or from indirect displacement effects operating through rising land and labor costs? Future research could leverage micro-firm data or industry-level data, employing matching methods or mediation models to precisely identify the specific channels of these negative pathways. This would enable a deeper understanding of the complex logic by which structural distortions impede innovation and provide more actionable micro-level evidence for resolving them through institutional reform and digital governance.

Third, there remains considerable scope for expanding the contingency boundaries of DEG theory. Although this study identifies the constraining roles of regional economic development levels and city types on the innovation effect of digital government through moderation and heterogeneity analyses, the efficacy of DEG as a macro-governance paradigm may be influenced by a broader array of micro-institutional environments and governance structures. Future research should further examine contextual variables such as intergovernmental competition and cooperation dynamics, the level of data infrastructure development, and public digital literacy. Constructing more complex models would allow for a more comprehensive mapping of DEG theory's applicability across diverse

institutional settings, providing a finer-grained analytical framework for distinguishing between "effective digital governance" and "digital governance failure."

Author Contributions: Conceptualization, Z.Z. and D.W.; methodology, D.W.; software, D.W.; validation, Z.Z. and D.W.; formal analysis, Z.Z.; investigation, D.W.; resources, Z.Z.; data curation, D.W.; writing—original draft preparation, D.W.; writing—review and editing, Z.Z.; visualization, D.W.; supervision, Z.Z.; project administration, Z.Z.; funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The National Social Science Foundation of China (Grant No. 24AZD068).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available on request from the authors.

Acknowledgments: During the drafting of this manuscript/research, the authors utilized AI tools (The Grammarly Pro) specifically designed to refine English grammar and assist with translation. The authors have reviewed and edited the final version and assume full responsibility for the content of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Furman, J.L.; Porter, M.E.; Stern, S. The determinants of national innovative capacity. *Res. Policy* **2002**, *31*, 899–933. [CrossRef]
2. Porter, M. Competitive advantage of nations. *Compet. Intell. Rev.* **1990**, *1*, 14. [CrossRef]
3. Audretsch, D.B.; Feldman, M.P. R&D spillovers and the geography of innovation and production. *Am. Econ. Rev.* **1996**, *86*, 630–640.
4. North, D.C. *Institutions, Institutional Change and Economic Performance*; Cambridge University Press: Cambridge, UK, 1990.
5. Lundvall, B. National Innovation Systems—Analytical Concept and Development Tool. *Ind. Innov.* **2007**, *14*, 95–119. [CrossRef]
6. Janowski, T. Digital government evolution: From transformation to contextualization. *Gov. Inf. Q.* **2015**, *32*, 221–236. [CrossRef]
7. Mergel, I.; Rethemeyer, R.K.; Isett, K. Big Data in Public Affairs. *Public. Adm. Rev.* **2016**, *76*, 928–937. [CrossRef]
8. Gil-Garcia, J.R.; Dawes, S.S.; Pardo, T.A. Digital government and public management research: Finding the crossroads. *Public. Manag. Rev.* **2018**, *20*, 633–646. [CrossRef]
9. Zhang, J.; Li, Y.; Liu, N.; You, J. Does digital government promote collaborative innovation? Evidence from the e-government pilot policy in China. *PLoS ONE* **2025**, *20*, e0334131. [CrossRef] [PubMed]
10. Li, L.; Peng, J.; Hu, B. Bridging the Regional Innovation Gap: The Driving Role of Digital Governance in China. *Growth Change* **2026**, *57*, e70098. [CrossRef]
11. Williamson, O.E. *The Economic Institutions of Capitalism*; Free Press: New York, NY, USA, 1985.
12. Janssen, M.; Charalabidis, Y.; Zuiderwijk, A. Benefits, adoption barriers and myths of open data and open government. *Inf. Syst. Manag.* **2012**, *29*, 258–268. [CrossRef]
13. Barney, J. Firm resources and sustained competitive advantage. *J. Manag.* **1991**, *17*, 99–120. [CrossRef]
14. Huang, Y.; Ye, D.; Xu, S. Optimal industrial structure and economic growth: A new measurement method and application. *Kybernetes* **2025**, *54*, 502–521. [CrossRef]
15. Lin, J.Y. *New Structural Economics: A Framework for Rethinking Development and Policy*; The World Bank: Washington, DC, USA, 2012.
16. DiMaggio, P.J.; Powell, W.W. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *Am. Sociol. Rev.* **1983**, *48*, 147–160. [CrossRef]
17. Scott, W.R. *Institutions and Organizations: Ideas and Interests*, 3rd ed.; Sage Publications: Thousand Oaks, CA, USA, 2008.
18. Tornatzky, L.G.; Fleischer, M.; Chakrabarti, A.K. *Processes of Technological Innovation*; Lexington Books: Lanham, MD, USA, 1990.
19. Zhu, K.; Kraemer, K.L.; Xu, S. The process of innovation assimilation by firms in different countries: A technology diffusion perspective on e-business. *Manag. Sci.* **2006**, *52*, 1557–1576. [CrossRef]
20. Dunleavy, P.; Margetts, H.; Bastow, S.; Tinkler, J. New public management is dead—Long live digital-era governance. *J. Public Adm. Res. Theory* **2006**, *16*, 467–494. [CrossRef]
21. Tan, X.; Li, Y.; Zhao, S. Digital government, transaction costs, and green innovation: Evidence from China. *Econ. Anal. Policy* **2024**, *81*, 112–130. [CrossRef]
22. Layne, K.; Lee, J. Developing fully functional E-government: A four stage model. *Gov. Inf. Q.* **2001**, *18*, 122–136. [CrossRef]

23. Cooke, P. Regional innovation systems: Competitive regulation in the new Europe. *Geoforum* **1992**, *23*, 365–382. [CrossRef]
24. Asheim, B.T.; Isaksen, A. Location, agglomeration and innovation: Regional innovation systems in Norway. *Eur. Plan. Stud.* **1997**, *5*, 299–330. [CrossRef]
25. Li, H.; Chen, X. Digital technology administrative penalties and green technology innovation: Evidence from China. *Heliyon* **2024**, *11*, e41159. [CrossRef] [PubMed]
26. Hao, X.; Miao, E.; Sun, Q.; Li, K.; Wen, S. The impact of digital government on corporate green innovation. *Technol. Forecast. Social. Change* **2024**, *206*, 123570. [CrossRef]
27. Marshall, A. *Principles of Economics*; Macmillan and Co: New York, NY, USA, 1890.
28. Kalathil, N.; Lanahan, L.; Feldman, M.; Fuchs, E.R.H. Varieties of agglomeration: Disentangling horizontal and vertical agglomeration within the manufacturing sector in the United States. *Res. Policy* **2025**, *54*, 105272. [CrossRef]
29. Lu, D.; Hui, E.C.M.; Shen, J.; Shi, J. Digital industry agglomeration and urban innovation: Evidence from China. *Econ. Anal. Policy* **2024**, *84*, 1998–2025. [CrossRef]
30. Martins, J.; Veiga, L.G. Digital government as a business facilitator. *Inf. Econ. Policy* **2022**, *60*, 100990. [CrossRef]
31. Yang, Y.; Zhu, Y. Digital industry agglomeration and firm innovation: Theory and evidence. *Manag. Decis. Econ.* **2026**, *47*, 357–386. [CrossRef]
32. Hsieh, C.-T.; Klenow, P.J. Misallocation and manufacturing TFP in China and India. *Q. J. Econ.* **2009**, *124*, 1403–1448. [CrossRef]
33. Bai, J.H.; Bian, Y.Q. Factor market distortion and the efficiency losses of Chinese innovative production. *China Ind. Econ.* **2016**, 139–154. (In Chinese) [CrossRef]
34. Brandt, L.; Van Biesebroeck, J.; Zhang, Y. Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing. *J. Dev. Econ.* **2012**, *97*, 339–351. [CrossRef]
35. Ren, S.; Wei, W. The impact of digital government on innovation: City-level evidence from China. *Emerg. Mark. Financ. Trade* **2026**, *62*, 689–703. [CrossRef]
36. Zhuo, Y.D.; Wang, Q. Digital infrastructure and innovation: Digital divide or digital dividend? *J. Innov. Knowl.* **2024**, *9*, 100542. [CrossRef]
37. Goldfarb, A.; Tucker, C. Digital Economics. *J. Econ. Lit.* **2019**, *57*, 3–43. [CrossRef]
38. Faggian, A.; McCann, P. Human capital, graduate migration and innovation in British regions. *Camb. J. Econ.* **2009**, *33*, 317–333. [CrossRef]
39. Hagedoorn, J.; Cloodt, M. Measuring innovative performance: Is there an advantage in using multiple indicators? *Res. Policy* **2003**, *32*, 1365–1379. [CrossRef]
40. Cohen, W.M.; Levinthal, D.A. Absorptive capacity: A new perspective on learning and innovation. *Adm. Sci. Q.* **1990**, *35*, 128–152. [CrossRef]
41. Jiao, J.; Jinhua, G.; Guohao, Z. Digital Industry Agglomeration, Local Government Competition and Urban Green Economy Efficiency Improvement. *Econ. Surv.* **2023**, *40*, 51–60. (In Chinese) [CrossRef]
42. Yue, X.; Shi, Q.W. Digital economy, innovation capability, and new quality productivity. *Financ. Res. Lett.* **2025**, *88*, 109112. [CrossRef]
43. Tang, Z.W.; Li, X.Y. (Eds.) *Report on the Development of Internet Service Capability of Local Governments in China (2019)*; Institute of Smart Governance, University of Electronic Science and Technology of China/Social Sciences Academic Press: Beijing, China, 2019.
44. Liu, H.; Shi, Y.; Sun, B.; Yang, X. Agglomeration of the Digital Services Industry and Digital Transformation: Evidence from China. *Emerg. Mark. Financ. Trade* **2024**, *60*, 855–869. [CrossRef]
45. Li, J.; Sutherland, D.; Ning, L.; Wang, Y. Firm ownership, industrial structure, and regional innovation performance in China's provinces. *Technol. Anal. Strateg. Manag.* **2014**, *26*, 1001–1022. [CrossRef]
46. Gao, K.; Yuan, Y. Government intervention, spillover effect and urban innovation performance: Empirical evidence from national innovative city pilot policy in China. *Technol. Soc.* **2022**, *70*, 102035. [CrossRef]
47. Ma, D.; Zhu, Q. Innovation in emerging economies: Research on the digital economy driving high-quality green development. *J. Bus. Res.* **2022**, *145*, 801–813. [CrossRef]
48. Allen, F.; Qian, J.; Qian, M. Law, finance, and economic growth in China. *J. Financ. Econ.* **2005**, *77*, 57–116. [CrossRef]
49. Baron, R.M.; Kenny, D.A. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Personal. Social. Psychol.* **1986**, *51*, 1173–1182. [CrossRef] [PubMed]
50. Cecchetti, S.G.; Kharroubi, E. Reassessing the impact of finance on growth. In *BIS Working Papers, No. 381*; Bank for International Settlements: Basel, Switzerland, 2012. Available online: https://www.bis.org/publ/work381.htm (accessed on 23 June 2026).
51. Griliches, Z. Patent statistics as economic indicators: A survey. *J. Econ. Lit.* **1990**, *28*, 1661–1707. Available online: http://www.jstor.org/stable/2727442 (accessed on 23 June 2026).
52. Acs, Z.J.; Audretsch, D.B. Patents as a measure of innovative activity. *Kyklos* **1989**, *42*, 221–234. [CrossRef]

53. Arundel, A.; Kabla, I. What percentage of innovations are patented? Evidence from German manufacturing industries. *Res. Policy* **1998**, *27*, 127–145. [CrossRef]
54. Pakes, A.; Griliches, Z. Estimating distributed lags in short panels with an application to the specification of depreciated capital and the patent–R&D relationship. *Econometrica* **1984**, *52*, 909–938. [CrossRef]
55. Hall, B.H.; Jaffe, A.; Trajtenberg, M. Market Value and Patent Citations. *RAND J. Econ.* **2005**, *36*, 16–38. Available online: http://www.jstor.org/stable/1593752 (accessed on 23 June 2026).
56. OECD/Eurostat. *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation*, 4th ed.; OECD Publishing: Paris, France, 2018. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

healthcare

MDPI

Article

Identifying Key Predictors of Nursing Workload in Emergency Infusion Rooms: A Decision Tree Approach

Leiming Gao, Ruixin Shi, Liuzi Wang, Shengzhi Jiao and Bei Wang *

School of Nursing, Nanjing University of Chinese Medicine, Nanjing 210023, China
* Correspondence: wwthk1998@njucm.edu.cn

Abstract

Purpose: Accurate assessment of nursing workload is essential for staffing allocation and operational management in emergency infusion rooms. However, workload generation is influenced by complex and potentially nonlinear interactions among patient volume, treatment duration, and care activities, which may not be adequately captured by conventional statistical approaches. This study aimed to identify key predictors associated with nursing workload intensity and develop an interpretable workload stratification framework using a Classification and Regression Tree (CRT) model. **Methods:** Daily operational data were collected from an emergency infusion room between July 2023 and August 2025. Daily chair utilization rate was used as a proxy indicator of workload intensity. Candidate predictors included total infusion duration, direct care encounters, number of patients receiving infusions, medication dispensing time, severe dependency, fall-risk patients, and triage-level patient volume. A CRT model was developed to identify hierarchical predictor relationships and threshold-based workload classification rules. Model robustness was evaluated using 10-fold cross-validation, comparative analyses with multiple linear regression, random forest, and gradient boosting models, and sensitivity analyses excluding total infusion duration. **Results:** The analysis included 761 valid observation days. Total infusion duration emerged as the most influential predictor, followed by direct care encounters and the number of patients receiving infusions. The CRT model identified clinically interpretable workload thresholds and generated a parsimonious decision structure for workload stratification. Re-substitution and cross-validation risk estimates were 0.045 (SE = 0.005) and 0.046 (SE = 0.005), respectively, indicating stable model performance. Although random forest and gradient boosting achieved higher predictive accuracy, the CRT model provided greater interpretability through transparent decision rules. Sensitivity analyses demonstrated that the overall workload stratification pattern remained largely unchanged after excluding total infusion duration. **Conclusions:** The CRT model identified total infusion duration, direct care encounters, and patient volume as key predictors associated with workload intensity in emergency infusion rooms. Although alternative models achieved higher predictive performance, the CRT approach provided interpretable workload stratification rules that may support staffing allocation and operational decision-making. The findings offer a practical data-driven framework for workload assessment in infusion care settings.

Keywords: nursing workload; decision tree approach; emergency department

1. Introduction

Nursing workload comprises the time, physical effort, and cognitive demand required for nursing activities, including technical complexity and associated risks [1]. Nursing workload is a multidimensional construct encompassing direct patient care, indirect nursing activities, documentation responsibilities, and organizational tasks, all of which contribute to overall nursing demand and resource utilization [2–4]. The infusion room is a critical outpatient service unit, which presents unique challenges due to its distinct workflow compared with inpatient wards, limiting the applicability of conventional ward-based workload evaluations.

Existing nursing workload evaluation approaches, including task-based measurement systems and technology-assisted monitoring tools, may still inadequately capture the complexity and dynamic variability of clinical nursing demands [5]. Recent studies have increasingly emphasized the integration of nursing time, care complexity, and technology-assisted workload assessment to improve the accuracy and responsiveness of workload evaluation systems [1,6]. However, these approaches lack adaptability to dynamic clinical environments, such as infusion rooms. The emergency infusion room is a high-intensity clinical setting where nursing workload is influenced by a complex interplay of factors including patient criticality, infusion type, patient turnover rate, and indirect nursing activities. Accurate, real-time assessment of this workload is crucial for effectively allocating nursing staff, reducing occupational burnout, and maintaining nursing quality.

Although nursing workload is a multidimensional construct, direct measurement of overall workload remains challenging in dynamic infusion settings. Therefore, daily chair utilization rate was adopted as a proxy indicator of workload intensity in the present study. Chair utilization rate is widely used as an operational metric for evaluating infusion center capacity and productivity because it reflects the proportion of occupied chair time relative to available chair capacity. Previous studies [7,8] have shown that chair utilization is closely associated with patient flow, treatment duration, staffing demand, and workflow efficiency in infusion services. Accordingly, although chair utilization does not directly measure nursing workload, it provides a practical and routinely available indicator of workload intensity and operational demand in emergency infusion rooms.

Previous studies [9,10] examining nursing workload have primarily relied on conventional statistical approaches, such as correlation analysis and linear regression. However, workload generation is often influenced by nonlinear relationships, complex interactions among operational factors, and hierarchical decision patterns that may not be adequately represented by traditional linear models. Recent studies [11–13] have suggested that machine-learning approaches may better capture the multidimensional and dynamic characteristics of healthcare workload and operational demand, particularly in complex clinical environments characterized by heterogeneous patient populations and fluctuating service requirements. Decision tree algorithms offer a promising alternative due to their ability to model nonlinear relationships, process multidimensional data, and generate intuitive classification rules directly applicable to clinical settings [14]. Recent studies have increasingly explored machine learning approaches for healthcare workload prediction and emergency resource optimization. For example, Song et al. applied machine learning models to predict nursing workload in digestive wards and demonstrated improved workload estimation performance [15]. Similarly, El Ariss et al. developed a machine-learning-based framework for emergency department resource allocation, highlighting the value of predictive analytics for staffing optimization [16]. Recent advances in explainable artificial intelligence have further emphasized the importance of interpretable prediction models in emergency care environments [17,18]. Interpretable machine-learning approaches allow clinicians and managers to understand the rationale underlying model predictions, thereby facilitating

https://doi.org/10.3390/healthcare14131966

trust, accountability, and practical implementation. In nursing management and healthcare operations, explainable models are particularly important because workforce allocation decisions often require clear justification and stakeholder acceptance [19–21]. However, despite these developments, evidence regarding interpretable machine learning approaches specifically tailored to emergency infusion room nursing workload evaluation remains limited. The predictive accuracy and hierarchical importance of key infusion-specific factors—such as infusion duration, patient dependency level, and triage level—remain to be explored. Decision trees have high efficiency in extracting key predictors and generating interpretable classification rules, and thus they are widely used in many fields, ranging from business modeling to medical diagnostics [22]. Their stratified structure prioritizes variables by predictive importance [23], providing a transparent framework for workload analysis.

Despite increasing interest in nursing workload management, limited evidence has explored the application of interpretable machine-learning approaches to workload assessment in emergency infusion rooms. Furthermore, few studies have established operationally applicable models capable of identifying key predictors and threshold-based workload patterns in this setting. Therefore, this study aimed to develop and validate a classification and regression tree (CRT) model to identify key predictors associated with nursing workload intensity, represented by daily chair utilization rate, in an emergency infusion room. In addition, we sought to establish interpretable workload stratification rules that may support staffing allocation and operational decision-making in clinical practice.

2. Literature Review

2.1. Nursing Workload in Emergency Settings

Nursing workload is a multidimensional construct involving direct patient care, indirect nursing activities, documentation, communication, technical procedures, and organizational tasks. Existing studies have developed and validated several nursing workload measurement tools, including intervention-based workload scales and multicenter workload assessment instruments, which provide important methodological foundations for quantifying nursing demand in clinical settings.

Previous studies examining nursing workload have commonly relied on conventional statistical approaches, including correlation analysis, linear regression, and multivariate regression models, to explore associations between patient characteristics, staffing factors, workflow variables, and perceived workload. For example, Ivziku et al. [10] identified patient acuity, staffing resources, patient transfers, documentation, isolation, and unscheduled activities as predictors of perceived nursing workload using regression models. Similarly, Griffiths et al. [24] demonstrated that nursing workload and staffing variation were significantly associated with patient outcomes, highlighting the complexity of workload generation and its implications for healthcare quality. These studies provide valuable evidence for understanding workload-related factors; however, they may not fully capture nonlinear relationships and interaction effects among multiple operational indicators.

In infusion-related settings, several studies have attempted to quantify nursing workload using technology-assisted or task-based approaches. For instance, Huiqing [25] used Personal Digital Assistants (PDAs) to quantify infusion-specific tasks (e.g., drug dispensing and venipuncture), and linked these metrics to performance appraisal, demonstrating enhanced nurse motivation and management quality. Juanjuan [26] identified daily patient volume and peak workload as critical indicators, showing that quantitative models reduce errors and increase stakeholder satisfaction. Based on this study, Junyu [27] used a Mobile Nursing Information System (MNIS) to measure the clinical workload (e.g., IV insertion and resuscitation), underscoring the value of informatics in achieving accurate

nurse-centric workload evaluations. However, both approaches share a common limitation: they primarily focus on direct, observable tasks (such as venipuncture and resuscitation), potentially overlooking indirect nursing activities (such as paperwork, family communication, and equipment handling) that constitute a significant portion of the overall nursing workload.

These studies reveal a shift toward data-driven workload metrics in infusion care. However, a lack of standardization persists, and existing methods often overlook the hierarchical relationships between workload predictors. The aforementioned tools share a common limitation: they struggle to capture the inherent complexity of infusion room workload. Three characteristics render traditional approaches inadequate:

Nonlinearity. Workload spikes triggered by external events—such as mass casualties—defy accurate prediction by linear models.

Interactive effects. Simultaneously managing critically dependent patients while preparing multiple high-risk infusions creates compound demands far exceeding simple addition.

Hierarchical heterogeneity. Patient needs vary dramatically within the same triage level—a Level 3 patient requiring a single antibiotic infusion versus multiple concurrent infusions presents vastly different workload demands.

These characteristics—nonlinearity, interaction effects, and heterogeneity—are consistent with the analytical strengths of decision tree methods. Decision trees recursively partition data based on significant predictors, capturing nonlinear relationships without requiring predefined functional forms. Their hierarchical structure mirrors patient severity stratification, while intuitive if-then rules provide the interpretability needed for clinical management decisions.

We further clarified that information systems and mobile nursing platforms were not discussed as technological innovations alone, but as tools for capturing workload-related data, quantifying nursing activities, and supporting objective workload assessment and performance evaluation. Recent advances in digital health technologies and machine learning-based workload monitoring systems have further expanded the possibilities for real-time nursing workload assessment and dynamic workforce management in high-intensity clinical settings [5,6].

2.2. Decision Tree Applications in Healthcare and Operational Management

Recent studies have increasingly applied machine-learning approaches to healthcare workload prediction, patient flow management, and resource allocation. Song et al. [15] used machine-learning algorithms to dynamically predict nursing workload in digestive wards based on patient characteristics, demonstrating the feasibility of machine-learning methods for nursing workload prediction. Moreno-Sánchez et al. [17] applied explainable artificial intelligence to predict patient flow in emergency departments, supporting the value of interpretable models in emergency resource allocation. Mafirakureva [28] modeled the cost-effectiveness of pediatric tuberculosis (TB) service integration in Africa, using decision trees to analyze outcomes, such as the Disability-Adjusted Life Year (DALY) and mortality; the authors reported that results depend on baseline service coverage. Tesfa [29] used this method to predict pregnancy termination factors in East Africa, reaching a model accuracy of 71.8%, which identified socioeconomic variables (e.g., education and marital status) as key influencers. Juntao [30] demonstrated the effectiveness of decision trees in surgical wards, where admission rates emerged as the primary workload predictor. These studies suggest that machine-learning methods may provide flexible tools for modeling complex healthcare processes involving multidimensional and dynamic clinical data.

Among machine-learning methods, decision tree algorithms are particularly useful when interpretability and operational applicability are important. Unlike black-box models, decision trees generate transparent threshold-based rules that can be directly understood by clinicians and nursing managers. This feature is especially valuable in nursing management, where staffing decisions require clear and explainable criteria. Comparative studies have shown that ensemble-learning methods such as Random Forest and Gradient Boosting often achieve superior predictive performance compared with single decision trees because they reduce variance and improve generalization [31]. However, these improvements are typically achieved at the expense of model transparency and interpretability. In healthcare settings, where model output may influence staffing decisions and operational management [32], interpretable models remain valuable because their decision pathways can be readily understood and validated by clinical stakeholders.

Decision tree models have been applied in various healthcare contexts, including clinical prediction, patient classification stratification, emergency department triage, and resource planning. However, limited evidence has explored the application of decision tree regression models to nursing workload stratification in emergency infusion rooms. More specifically, few studies have investigated whether routinely collected infusion-room indicators can be used to identify interpretable threshold rules for workload intensity. Therefore, applying a classification and regression tree model in this setting may help identify key predictors, clarify their hierarchical contributions, and provide operationally meaningful rules for nursing workload management.

Prior to initiating the study, a systematic literature search was conducted to ascertain the novelty of applying decision tree regression models to the emergency infusion room context. The search was performed across PubMed, CINAHL, Web of Science and the China National Knowledge Infrastructure (CNKI) databases, covering all records from inception through June 2025. The search strategy employed combinations of the following keywords: "decision tree," "regression tree," "CART," "nursing workload," "emergency department," "infusion room," and "outpatient infusion." The search results confirmed that while machine learning methods, including decision trees, have been increasingly applied to predict patient volumes and acuity in general emergency departments, limited peer-reviewed evidence was identified regarding the use of decision tree regression algorithms to model or predict nursing workload indicators within emergency infusion rooms. More precisely, this paper leverages the effectiveness of decision trees to identify dominant workload indicators (e.g., infusion duration and patient volume) and establish a transparent rule-based framework for resource allocation, aligning with value-based healthcare principles.

3. Methodology

3.1. Data Collection

This study retrospectively analyzed daily workload data from the infusion room of a Grade-A tertiary hospital in Jiangsu Province, China, covering the period from 1 July 2023 to 31 August 2025. The workload indicators comprise clinical activity metrics (i.e., the daily chair utilization rate, direct care encounters, number of patients receiving infusions, patient transfers, resuscitations, and deaths), patient complexity metrics (counts of high fall-risk patients, patient volume by triage levels, including Levels II, III, and IV, and patients with severe dependency), and time-based metrics (total daily infusion duration and medication dispensing time).

Daily chair utilization rate was calculated as: (total chair occupancy time/total available chair time) $\times$ 100%.

https://doi.org/10.3390/healthcare14131966

The data are provided by Hospital Information System (HIS) (automatically recorded metrics such as infused patients, direct care episodes, and dispensing time) and nurse manager reports (manually documented high-risk and high-dependency cases).

From an initial 792-day dataset, 31 days are excluded due to missing or incomplete records, yielding 761 valid days for analysis. Missing data were assessed using Little's MCAR test. The results indicated that the missing observations were Missing Completely at Random (MCAR) ($p > 0.05$). Because the proportion of missing data was relatively small and the MCAR assumption was satisfied, no imputation procedures were applied. Instead, listwise deletion was performed, resulting in the exclusion of 31 observation days with incomplete records. Under the MCAR assumption, listwise deletion provides unbiased parameter estimates and is considered an appropriate approach for handling missing data. The final analysis included 761 complete observation days. The excluded observation days were distributed throughout the study period and were not concentrated during specific seasons, public holidays, or unusually high-demand periods. Therefore, listwise deletion was unlikely to introduce meaningful seasonal or temporal bias into the final analytical dataset.

3.2. Data Preprocessing

The Classification and Regression Tree (CRT) algorithm in SPSS 29.0 is used to evaluate the nursing workload in the emergency infusion room and analyze the predictor importance and classification stratification rules. The dependent variable was daily chair utilization rate, which served as a proxy measure of nursing workload intensity in the emergency infusion room. This indicator was selected because it reflects the combined influence of patient volume, infusion duration, patient turnover, and operational demand on available nursing resources. The remaining workload metrics collected are used as independent variables.

The CRT analysis automatically identifies hierarchical relationships between variables, generating an optimized decision tree based on predictor significance.

For the regression tree predicting continuous nursing workload, the splitting criterion was based on variance reduction (also termed impurity reduction for regression), which measures the decrease in within-node variance achieved by partitioning a parent node into two child nodes.

Gain Calculation Formula: For a given split s at node t, the gain $\Delta(s,t)$ is computed as:

$$\Delta(s,t) = SS(t) - [SS(t_L) + SS(t_R)]$$

where:

- $SS(t) = \sum_{i \in t}(y_i - \overline{y}_t)^2$ is the sum of squared deviations from the mean in the parent node;
- $SS(t_L)$ and $SS(t_R)$ are the sums of squared deviations in the left and right child nodes, respectively.

The gain percentage reported for each node represents the proportion of the total variance explained by that split, normalized across all splits in the tree. Specifically, it is calculated as:

$$Gain\% = \frac{\Delta(s,t)}{SS_{total}} \times 100\%$$

where SS_{total} is the total sum of squares of the dependent variable.

To avoid potential misunderstanding, it should be noted that negative node means observed in the raw CRT output do not represent negative nursing workload or negative nursing time. These values arise from the internal scaling and calculation procedures

https://doi.org/10.3390/healthcare14131966

used during model construction and reflect relative deviations from the overall sample mean rather than actual workload levels. For presentation purposes, all workload values reported in this study were converted back to their original scale, ensuring that the results can be interpreted directly in terms of workload intensity.

3.3. Development of the Decision Tree Model

The CRT algorithm can be used for classification and regression. The output of a classification tree is the category of the sample, while that of a regression tree is a real number. It comprises three steps: feature selection, recursive construction of decision tree, and decision tree pruning.

The CRT model was implemented using binary recursive partitioning in SPSS. Surrogate splitting was not enabled because missing data were addressed prior to model development. Tree complexity was controlled through pre-pruning parameters, including a maximum depth of three levels, a minimum parent node size of 100 observations, and a minimum child node size of 50 observations. Cost-complexity post-pruning was not applied.

Maximum depth = 3: During cross-validation, tree depths ranging from 2 to 10 were evaluated. Depth 3 yielded the optimal trade-off between predictive performance (minimizing cross-validated MSE) and interpretability. Depths greater than 3 produced marginal improvements in accuracy ($\leq$2%) but generated substantially more complex rules, reducing clinical usability. Depth 2, while simpler, resulted in a 12% increase in prediction error. Thus, depth 3 was selected as the parsimonious optimum.

Minimum parent node size =100 and minimum child node size = 50: These parameters control the minimum number of cases required for a node to be split (parent) and for a terminal node (child). These values were optimized via grid search during cross-validation. The selected thresholds ensure that each split is based on a sufficiently large sample ($\geq$100 shifts) to yield stable, generalizable patterns, while preventing splits that create tiny, potentially overfitted terminal nodes (minimum 50 shifts). These sizes represent approximately 15% and 7.5% of the training sample, respectively, striking a balance between capturing meaningful subgroups and avoiding noise-driven splits.

3.4. Comparative Model Analysis

To benchmark predictive performance, multiple linear regression, random forest, and gradient boosting models were developed using the same predictor set and evaluated using 10-fold cross-validation. Model performance was compared using R^2, RMSE, MAE, and MSE.

4. Results

4.1. Descriptive Statistics of the Independent Variable

Table A1 shows the final inclusion of 761 days of infusion suite workload data, containing 10 representative workload indicators, along with descriptive data.

Output of the Decision Tree Model

The CRT algorithm generated a parsimonious tree with a maximum depth of three, yielding six terminal nodes that classify shift-level workload into low, medium, and high categories. Figure 1 presents the visual structure of the decision tree, with splitting rules and threshold values at each node.

Root Node Split: The total daily infusion duration emerged as the most important predictor, forming the initial partition of the data. Shifts with total infusion duration $\leq$ 305 h

https://doi.org/10.3390/healthcare14131966

were classified directly into the low workload stratum, representing periods with minimal nursing time requirements.

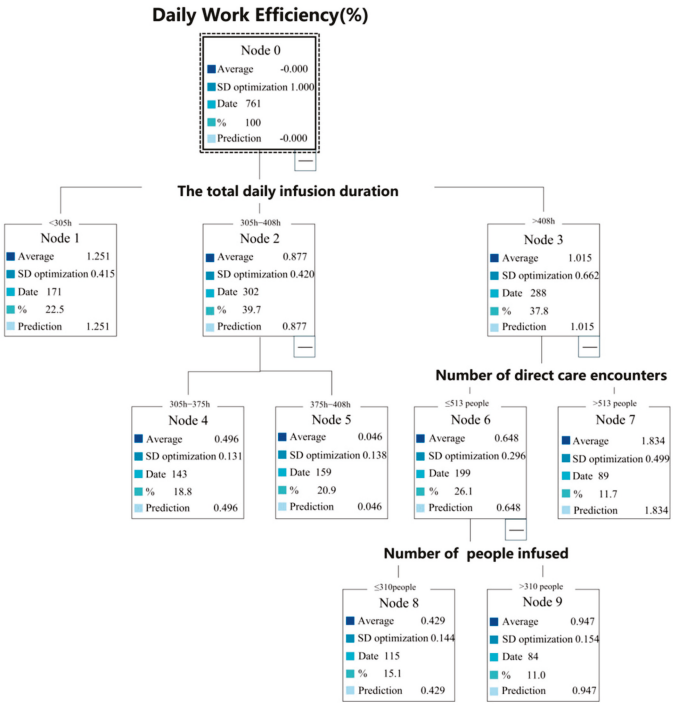


Figure 1. Decision Tree for Infusion Suite Workload Indicators Tree Chart.

Secondary Splits for Intermediate Infusion Durations: For shifts with infusion duration between 305 and 408 h, workload stratification required additional predictors. At this node, number of direct care encounters became the decisive factor: shifts exceeding 513 direct care encounters were classified as high workload, while those below this threshold proceeded to the next level.

Tertiary Split for Patient Volume: Among shifts with infusion duration in the intermediate range (305–408 h) and direct care encounters $\leq$ 513, number of patients receiving infusions provided the final differentiation. Shifts with patient volume > 310 were classified as high workload, whereas those with $\leq$310 patients were categorized as medium workload.

High-Duration Direct Classification Stratification: Shifts with total infusion duration > 408 h were classified directly as high workload without further splitting, indicating that extreme infusion duration alone is sufficient to predict peak nursing demand.

Figure 1 presents the CRT model illustrating the hierarchical structure of the decision tree.

Table 1 presents the gain statistics for each terminal node of the CRT model. In CRT analysis, the gain percentage reflects the proportion of the overall target outcome captured by a specific node relative to the entire dataset. Higher gain percentages indicate that a node contains a greater concentration of observations associated with higher workload intensity and therefore contributes more substantially to distinguishing workload patterns within the model.

Among the terminal nodes, Node 1 exhibited the highest gain percentage (22.5%) with a mean workload-intensity value of 1.251, indicating that this subgroup accounted for a relatively large share of the overall workload burden identified by the model. In contrast, Node 9 showed the lowest gain percentage (11.0%) with a mean value of 0.947, suggesting a comparatively smaller contribution to the overall workload distribution.

Table 2 presents risk evaluation metrics derived from a decision tree model, comparing two validation methods: *resubstituting* and *10-fold cross-validation*. The resubstituting error (0.045) represents the in-sample risk of the model, indicating how well the decision tree fits the training data.

Table 1. Summary of node gains.

Node	*n*	%	Mean Value
7	89	11.7%	1.834
9	84	11.0%	0.947
8	115	15.1%	0.429
5	159	20.9%	0.046
4	143	18.8%	0.496
1	171	22.5%	1.251

Table 2. Risk for the decision tree.

Method	Estimation	Standard Error
Re-substitute	0.045	0.005
Cross-validation	0.046	0.005

The close agreement between the re-substitution estimate (0.045, SE = 0.005) and cross-validation estimate (0.046, SE = 0.005) suggests stable predictive performance, supports the model's generalizability, and indicates limited evidence of overfitting.

To provide a more comprehensive assessment of model performance, additional regression metrics were calculated for the CRT model. The model achieved an R^2 of 0.939, indicating that approximately 93.9% of the variance in workload intensity was explained by the predictors included in the model. The corresponding MAE, MSE, and RMSE values were 0.035, 0.045, and 0.212, respectively, suggesting good predictive performance on the training dataset (Table 3).

Table 3. Performance metrics of the CRT model.

Metric	Training Set
R^2	0.939
MAE	0.035
MSE	0.045
RMSE	0.212

4.2. Importance of Variables

Figure 2 presents the relative importance of independent variables in the decision tree model, ranked by their contribution to the outcome prediction. The importance values are expressed as percentages, reflecting the impact of each variable on the decision-making process of the model. The highest importance score (0.922) indicates that total daily infusion duration was the most influential predictor within the fitted CRT model. Direct care encounters (0.804) and the number of patients receiving infusions (0.773) followed closely, suggesting that these variables contributed substantially to workload-intensity prediction within the model.

Medication dispensing time (0.634), severe dependency (0.622), and patient IV volume (0.621) demonstrated moderate predictive importance. Their similar importance scores suggest comparable contributions to the decision-tree splitting process.

Patient III (0.446), high risk of falling (0.300), and patient II (0.202) have a lesser effect. For example, the high risk of falling (32.5%) may affect classification stratification only in specific branches.

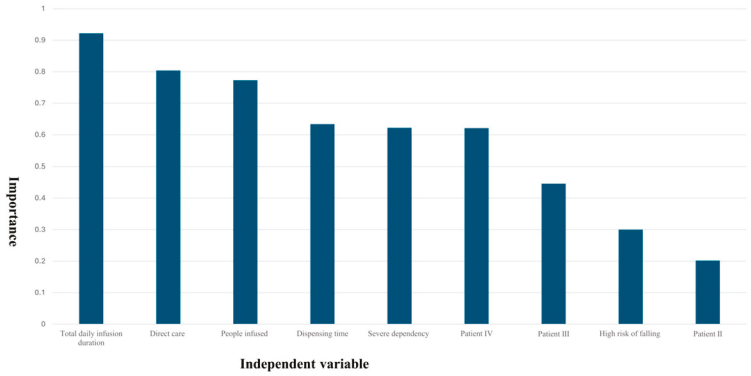


Figure 2. Importance of Independent variable.

4.3. Comparative Model Analysis

To benchmark the predictive performance of the CRT model, additional models including multiple linear regression, random forest, and gradient boosting were developed using the same predictor set and evaluated using 10-fold cross-validation. As shown in Table 4, alternative models achieved higher predictive accuracy than the CRT model. However, the CRT model was retained because it provides transparent threshold-based decision rules that can be readily translated into staffing and workload management strategies in emergency infusion rooms.

Table 4. Comparative predictive performance of different modeling approaches based on 10-fold cross-validation.

Model	R^2	RMSE	MAE
Linear Regression	1.000	<0.001	<0.001
Random Forest	0.993	0.015	0.008
Gradient Boosting	0.999	0.007	0.005
CRT	0.927	0.212	0.037

4.4. Sensitivity Analysis

To assess the potential influence of feature leakage between daily chair utilization rate and total infusion duration, a sensitivity analysis was conducted by excluding total infusion duration from the predictor set and rebuilding the CRT model using the same modeling parameters. As shown in Table 5, the sensitivity model demonstrated a reduction in predictive performance compared with the original model. The cross-validated R^2 decreased from 0.927 to 0.850, and the RMSE decreased from 0.212 to 0.073. Despite this decline in accuracy, the model retained acceptable predictive performance. Direct care encounters remained the most influential predictor, followed by the number of patients receiving infusions and medication dispensing time.

Table 5. Sensitivity analysis excluding total infusion duration.

Model	R^2	MSE	RMSE	MAE
Original CRT model	0.927	0.045	0.212	0.037
Excluding total infusion duration	0.850	0.039	0.073	0.053

Furthermore, the overall hierarchical structure of the decision tree remained largely consistent, indicating that the workload stratification patterns identified by the CRT model were not solely dependent on total infusion duration. These findings suggest that although total infusion duration contributed substantially to model performance, the main conclusions of the study remained robust after its exclusion.

5. Discussion

5.1. Interpretation of Key Predictors

The CRT model identified total infusion duration, direct care encounters, and the number of patients receiving infusions as the most influential predictors of workload intensity within the fitted model. These findings are clinically plausible because each indicator captures a distinct dimension of operational demand in emergency infusion rooms.

Total infusion duration reflects cumulative chair occupancy and sustained nursing supervision requirements. Longer infusion times increase the duration of patient monitoring, medication management, and clinical observation, thereby contributing to higher workload intensity. Direct care encounters represent the frequency of nurse–patient interactions and are closely associated with nursing labor input. Similarly, the number of patients receiving infusions reflects overall service volume and operational demand. Together, these predictors capture temporal, activity-based, and volume-related dimensions of workload generation.

The present findings are broadly consistent with previous nursing-workload studies reporting that patient volume, care intensity, and nursing activity levels are major contributors to workload demand. For example, Ivziku et al. [10] identified patient acuity, staffing requirements, and clinical activities as important workload predictors, while Griffiths et al. reported associations between workload variation and patient-care outcomes. Consistent with these studies, total infusion duration, direct care encounters, and patient volume emerged as the most influential predictors within the present model. However, unlike conventional regression-based approaches that primarily estimate associations, the CRT model identified operational thresholds that stratified workload intensity into distinct workload strata, providing a more interpretable framework for management decision-making.

The identified thresholds provide operationally meaningful workload strata that may facilitate workload monitoring and staffing decisions.

5.2. Implications for Nursing Management

One important advantage of the CRT model is its ability to generate transparent workload stratification rules [24]. Unlike conventional regression models that primarily quantify statistical associations, the decision tree identifies operational thresholds that can be readily interpreted and translated into management actions.

For example, the CRT model identified specific thresholds for total infusion duration and direct care encounters that distinguished low-, moderate-, and high-workload strata. Because these indicators are routinely collected during daily operations, they could potentially be incorporated into workload-monitoring systems without requiring additional data collection. When workload indicators approach or exceed predefined thresholds, nurse managers may anticipate periods of increased workload demand and consider temporary staffing adjustments, task redistribution, or workforce reinforcement. Conversely, during lower-demand periods, staffing resources may be allocated more efficiently while maintaining patient-care quality.

In addition, the workload stratification framework may provide an objective reference for interpreting differences in workload exposure when evaluating staff performance

https://doi.org/10.3390/healthcare14131966

and resource utilization. Rather than relying solely on patient volume or staffing ratios, managers may use workload strata to contextualize operational demands across different shifts or observation periods.

It should be noted that the present study did not directly evaluate patient outcomes, nursing burnout, staffing efficiency, or the effectiveness of workload-informed staffing interventions. Therefore, the potential management benefits described above should be interpreted as practical applications of the proposed framework rather than outcomes demonstrated by the current study. Future prospective studies are needed to evaluate whether integrating workload stratification into routine management processes improves staffing decisions and operational performance.

5.3. Comparison with Previous Studies and Alternative Models

The findings of this study are consistent with previous research indicating that workload intensity is influenced by patient volume, care complexity, and nursing activity levels. Similar predictors have been identified in studies examining emergency department operations, nursing workload measurement, and healthcare resource utilization [33].

To further evaluate model performance, multiple linear regression, random forest, and gradient boosting models were developed using the same predictor set. Although these alternative models achieved higher predictive accuracy than the CRT model, the observed differences in prediction error were relatively modest and did not substantially alter the overall workload stratification patterns identified in this study. From a nursing management perspective, the primary objective of the proposed framework is to support workload categorization and staffing allocation rather than to generate highly precise numerical predictions. Therefore, the improved interpretability and explicit threshold-based decision rules provided by the CRT model may outweigh the modest reduction in predictive accuracy for routine operational decision-making. Nevertheless, ensemble-learning approaches combined with explainable artificial intelligence techniques, such as SHAP or LIME, may further improve predictive performance while preserving interpretability and should be explored in future research.

5.4. Strengths and Limitations

This study has several strengths. The use of an interpretable machine-learning approach enabled the identification of workload-related patterns within a real-world emergency infusion setting. In addition, the CRT model generated transparent workload stratification rules that may facilitate operational decision-making and staffing management. The robustness of the findings was further evaluated through comparative analyses and sensitivity analyses.

Several limitations should also be acknowledged. The study was conducted in a single tertiary hospital, which may limit the generalizability of the findings to other healthcare settings. Furthermore, daily chair utilization rate was used as a proxy measure of workload intensity rather than a direct measure of nursing workload. Although this indicator reflects multiple dimensions of operational demand, it may not fully capture all aspects of nursing activities.

A further limitation is that the CRT model was developed using a predefined set of routinely collected operational indicators. Although these variables captured important dimensions of workload demand, additional factors that may influence nursing workload, such as staffing composition, nurse experience, patient-acuity characteristics, workflow interruptions, and organizational factors, were not available for analysis. Consequently, the identified workload stratification rules should be interpreted within the context of the variables examined in this study.

https://doi.org/10.3390/healthcare14131966

In addition, only internal validation was performed in the present study. Finally, this was a retrospective single-center study conducted in one tertiary emergency infusion room. Therefore, the identified workload predictors, decision thresholds, and workload stratification rules should be interpreted within the context of the present study setting. External validation using multicenter datasets and alternative workload measures is warranted before broader clinical implementation can be recommended.

6. Conclusions

This study developed and evaluated an interpretable classification and regression tree (CRT) model to assess workload intensity in an emergency infusion room using routinely collected operational indicators. Within the fitted model, total infusion duration, direct care encounters, and the number of patients receiving infusions emerged as the most influential predictors of workload intensity, represented by daily chair utilization rate.

The CRT model identified transparent threshold-based workload stratification patterns and demonstrated acceptable predictive performance. Comparative analyses showed that alternative models, including multiple linear regression, random forest, and gradient boosting, achieved higher predictive accuracy. However, the CRT model provided greater interpretability and generated explicit decision rules that may be more readily translated into nursing management practice.

Sensitivity analyses further suggested that the overall workload stratification pattern remained relatively stable after excluding total infusion duration, supporting the robustness of the main findings. Nevertheless, the use of daily chair utilization rate as a proxy measure of workload intensity and the potential conceptual overlap with infusion duration should be considered when interpreting the results.

Overall, the proposed framework provides a practical and interpretable approach for workload assessment in emergency infusion settings and may support data-informed staffing allocation and operational decision-making. Future studies should validate the model across multiple institutions and explore alternative workload outcome measures to further enhance generalizability and applicability.

Author Contributions: Conceptualization, L.G.; Methodology, L.G.; Software, L.W. and S.J.; Formal analysis, R.S.; Investigation, R.S., L.W. and S.J.; Data curation, L.G.; Writing—original draft, L.G.; Writing—review & editing, B.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Jiangsu Province Graduate Student Practical Innovation Program 2025 under Grant [number SJCX25-1105].

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Research Department at Jiangsu Provincial Hospital of Integrated Traditional Chinese and Western Medicine (30 October 2025 No. 2025-LWKYZ-052). Written informed consent was obtained from all participants after providing a full explanation of the study procedures, potential risks, and benefits, in accordance with ethical standards. The study was conducted as a retrospective analysis of routinely collected operational data. No patient intervention or prospective data collection was performed. All data were anonymized prior to analysis.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: The authors gratefully acknowledge their colleagues at Jiangsu Provincial Hospital of Integrated Traditional Chinese and Western Medicine for their insightful discussions and technical assistance.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the findings reported in this study.

Appendix A

Table A1. Descriptive statistics.

	N	Scope	Minimum Value	Maximum Value	Value		Standard Deviation	Variance	Skewness		Kurtosis	
Independent Variable	**Statistic**	**Statistic**	**Statistic**	**Statistic**	**Statistic**	**Standard Error**	**Statistic**	**Statistic**	**Statistic**	**Standard Error**	**Statistic**	**Standard Error**
Total daily infusion duration	761	35.85	1.10	36.95	17.3163	0.23258	6.41611	41.166	0.365	0.089	0.197	0.177
Dispensing time	761	6.87	0.03	6.90	1.5623	1.10570	1.10570	1.223	1.749	0.089	3.435	0.177
People infused	761	354	13	367	183.98	59.542	59.542	3545.290	0.118	0.089	0.168	0.177
Direct care	761	1055	39	1094	559.99	184.904	184.904	34,189.531	0.138	0.089	0.226	0.177
Patient II	761	15	0	15	4.01	2.917	2.917	8.507	1.157	0.089	1.447	0.177
Patient III	761	98	2	100	35.79	18.699	18.699	349.665	0.703	0.089	0.142	0.177
Patient IV	761	294	11	305	144.12	46.429	46.429	2155.688	0.342	0.089	0.175	0.177
High risk of falling	761	40	0	40	10.51	6.557	6.557	43.000	1.209	0.089	1.626	0.177
Severe dependency	761	196	5	200	75.68	34.200	34.200	1169.612	0.536	0.089	0.133	0.177
Effective cases	761											

References

1. Swiger, P.A.; Vance, D.E.; Patrician, P.A. Nursing workload in the acute-care setting: A concept analysis of nursing workload. *Nurs. Outlook* **2016**, *64*, 244–254. [CrossRef] [PubMed]
2. Hellín Gil, M.F.; Roldán Valcárcel, M.D.; Seva Llor, A.M.; Ibáñez-López, F.J.; Mikla, M.; Montesinos, M.J.L. Validation of a nursing workload measurement scale based on the classification of nursing interventions for adult hospitalization units. *Int. J. Environ. Res. Public Health* **2022**, *19*, 15528. [CrossRef] [PubMed]
3. De Groot, K.; De Veer, A.J.E.; Munster, A.M.; Francke, A.L.; Paans, W. Nursing documentation and its relationship with perceived nursing workload: A mixed-methods study among community nurses. *BMC Nurs.* **2022**, *21*, 34. [CrossRef] [PubMed]
4. Hellín Gil, M.F.; Mikla, M.; Seva Llor, A.M.; Valcárcel, M.D.R.; Ibáñez-López, F.J.; Montesinos, M.J.L. Multicenter application of a nursing workload measurement scale in adult hospitalization units. *Int. J. Nurs. Sci.* **2022**, *9*, 460–466. [CrossRef] [PubMed]
5. Mohammadnejad, F.; Freeman, S.; Klassen-Ross, T.; Hemingway, D.; Banner, D. Impacts of technology use on the workload of registered nurses: A scoping review. *SAGE Open Nurs.* **2023**, *9*, 23779608231180189. [CrossRef] [PubMed]
6. Sánchez-Sánchez, M.M.; Campos-Asensio, C.; Arias-Rivera, S. Workloads of intensive care nurses: Validity of their estimation using mobile applications and comparison with Nursing Activities Score. *Enfermería Intensiv.* **2024**, *35*, 229–241. [CrossRef] [PubMed]
7. Home, J.H.C.A.; Zhao, J.; Sherman, B.; McCracken, A.; Watkins, S.; Rusinko, K. Development of Productivity Standards for Ambulatory Infusion Suite Nurses Within a Multi-Entity Health System. *Infus. J.* **2023**, *2*, 6–15. [CrossRef]
8. Milo, A. How Much Room Is Left? Understanding Infusion Capacity by Defining Utilization. *Oncol. Pract. Manag.* **2020**, *10*, 20–24.
9. LeanTaaS. Scheduling Tool Improves Infusion Center Efficiency. Targeted Oncology, 25 October 2017. Available online: https://www.targetedonc.com/view/scheduling-tool-improves-infusion-center-efficiency (accessed on 26 June 2026).
10. Ivziku, D.; Clari, M.; Piredda, M.; Matarese, M.; Messina, M.P.; Albanesi, B.; De Marinis, M.G. Defining Nursing Workload Predictors: A Pilot Study. *J. Nurs. Manag.* **2022**, *30*, 473–481. [CrossRef] [PubMed]
11. Maghsoud, F.; Sadeghi, M.; Hosseinzadeh, M.; Peyrovi, H. The Mediating Role of Implicit Rationing of Nursing Care, Job Satisfaction, and Emotional Exhaustion in the Relationship between Workload and Quality of Nursing Care. *BMC Nurs.* **2022**, *21*, 333. [CrossRef] [PubMed]
12. Sendak, M.P.; D'Arcy, J.; Kashyap, S.; Gao, M.; Nichols, M.; Corey, K.; Ratliff, W.; Balu, S. A Path for Translation of Machine Learning Products into Healthcare Delivery. *EMJ Innov.* **2020**, *4*, 44–52.

13. Mahmoudi, E.; Kamdar, N.; Kim, N.; Gonzales, G.; Singh, K.; Waljee, A.K. Use of Machine Learning in Predicting Healthcare Utilization. *Health Serv. Res.* **2020**, *55*, 124–133.
14. Kang, J.; Kwon, S.S.; Lee, Y. Clinical nurses' work-life balance prediction due to patient safety incidents using classification and regression tree analysis: A secondary data analysis. *BMC Nurs.* **2024**, *23*, 70. [CrossRef] [PubMed]
15. Song, Y.; Zhang, X.; Luo, D.; Shi, J.; Zang, Q.; Wang, Y.; Yin, H.; Xu, G.; Bai, Y. Predicting nursing workload in digestive wards based on machine learning: A prospective study. *BMC Nurs.* **2024**, *23*, 908. [CrossRef] [PubMed]
16. El Ariss, A.B.; Kijpaisalratana, N.; Ahmed, S.; Yuan, J.; Coleska, A.; Marshall, A.; Luo, A.D.; He, S. Development and validation of a machine learning framework for improved resource allocation in the emergency department. *Am. J. Emerg. Med.* **2024**, *84*, 141–148. [CrossRef] [PubMed]
17. Moreno-Sánchez, P.A.; Aalto, M.; van Gils, M. Prediction of patient flow in the emergency department using explainable artificial intelligence. *Digit. Health* **2024**, *10*, 1–18. [CrossRef]
18. Hu, Y.; Chan, C.W.; Dong, J.; Kazekjian, A.; Ophaswongse, C.; Sugalski, G.; Underwood, J.P.; Perotte, R. Implementing a prediction driven framework for emergency department nurse staffing to optimize real time decisions. *npj Health Syst.* **2025**, *2*, 16. [CrossRef]
19. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Infin. Fusion* **2020**, *58*, 82–115. [CrossRef]
20. Amann, J.; Blasimme, A.; Vayena, E.; Frey, D.; Madai, V.I. Explainability for artificial intel-ligence in healthcare: A multidisciplinary perspective. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 310. [CrossRef] [PubMed]
21. Tjoa, E.; Guan, C. A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4793–4813. [CrossRef] [PubMed]
22. Porto, B.M. Improving triage performance in emergency departments using machine learning and natural language processing: A systematic review. *BMC Emerg. Med.* **2024**, *24*, 219. [CrossRef] [PubMed]
23. Tohka, J.; van Gils, M. Evaluation of machine learning algorithms for health and wellness applications: A tutorial. *Comput. Biol. Med.* **2021**, *132*, 104324. [CrossRef] [PubMed]
24. Griffiths, P.; Saville, C.; Ball, J.E.; Jones, J.; Pattison, N.; Monks, T. Nursing workload, nurse staffing methodologies and tools: A systematic scoping review and discussion. *Int. J. Nurs. Stud.* **2020**, *103*, 103487. [CrossRef] [PubMed]
25. Lin, H.; Wen, Y.; Liu, T.; Liao, S.; Ma, W. Application of performance appraisal in infusion room using PDA as a platform. *China Med. Her.* **2017**, *14*, 161–164.
26. Su, J. Application effect of quantitative assessment in nursing management of infusion room in emergency department. *J. Chin. Med. Manag.* **2017**, *25*, 84–85. [CrossRef]
27. Gao, J.; Cao, Y.; He, B.; Hao, W. Construction and application of MNIS combined with weight coefficients as a nursing workload measurement system in infusion room. *China Clin. Res.* **2021**, *34*, 141–144. [CrossRef]
28. Mafirakureva, N.; Denoeud-Ndam, L.; Tchounga, B.K.; Otieno-Masaba, R.; Herrera, N.; Mukherjee, S.; Casenghi, M.; Tiam, A.; Dodd, P.J.; INPUT Study Group. Cost-effectiveness of integrating paediatric tuberculosis services into child healthcare services in Africa: A modelling analysis of a cluster-randomised trial. *BMJ Glob. Health* **2024**, *9*, e016416. [CrossRef] [PubMed]
29. Tesfa, G.A.; Demeke, A.D.; Seboka, B.T.; Tebeje, T.M.; Kasaye, M.D.; Gebremeskele, B.T.; Hailegebreal, S.; Ngusie, H.S. Employing machine learning models to predict pregnancy termination among adolescent and young women aged 15–24 years in East Africa. *Sci. Rep.* **2024**, *14*, 30047. [CrossRef] [PubMed]
30. Chi, J.-T.; Song, Z.-L.; Zhu, Y.-J.; Xie, J.-M.; Wang, X.-N.; Zhou, L. Application of decision tree classification method in nursing workload evaluation in surgical wards. *Chin. J. Pract. Nurs.* **2019**, *35*, 2148–2151.
31. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [CrossRef] [PubMed]
32. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef] [PubMed]
33. Morris, R.; MacNeela, P.; Scott, A.; Treacy, P.; Hyde, A. Reconsidering the conceptualization of nursing workload: A literature review. *J. Adv. Nurs.* **2021**, *77*, 4451–4463.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

applied sciences

MDPI

Article

Machine Learning Analysis of Scientific Projects for Decision Support: Empirical Evidence from CORDIS FP6

Arailym Medetbek [1,*], Aidos Mukhatayev [2,*], Svitlana Biloshchytska [3,4], Andrii Biloshchytskyi [3] and Dias Berlibek [5]

[1] School of Software Engineering, Astana IT University, Astana 010000, Kazakhstan
[2] School of General Educational Disciplines, Astana IT University, Astana 010000, Kazakhstan
[3] Department of Computational and Data Science, Astana IT University, Astana 010000, Kazakhstan; bsv@astanait.edu.kz (S.B.); a.b@astanait.edu.kz (A.B.)
[4] Department of Information Technologies, Kyiv National University of Construction and Architecture, 03680 Kyiv, Ukraine
[5] Information Security System, L.N. Gumilyov Eurasian National University, Astana 010000, Kazakhstan; 021201550905@enu.kz
* Correspondence: arailym.medetbek@astanait.edu.kz (A.M.); aidos.mukhatayev@astanait.edu.kz (A.M.); Tel.: +7-707-546-9769 (A.M.); +7-702-631-2244 (A.M.)

Abstract

The increasing complexity of scientific research projects and growing competition for public funding require data-driven approaches to support evidence-based decision-making in research management. This paper presents an integrated machine learning-based analytical pipeline for scientific project analysis, positioning the contribution as methodological integration and empirical insight rather than the development of new algorithms. Using open data from the CORDIS database on projects funded under the Sixth Framework Programme (FP6), the study combines three complementary analytical tasks—structure identification, budget prediction, and funding-completeness classification—within a single empirical decision-support workflow. Unsupervised analysis based on dimensionality reduction and clustering reveals distinct structural patterns in project characteristics and identifies atypical large-scale projects characterized by substantially higher budgets and consortium sizes. Regression models predict total project costs, achieving $R^2 \approx 0.50$ on unseen data; the gap between explained and unexplained variance is interpreted in relation to latent contextual factors not available in administrative records. Classification models distinguish between fully and partially funded projects (threshold $r \geq 0.70$); the high discriminative performance (AUC > 0.97) is interpreted cautiously in light of target-variable construction and potential feature dependence. Association rule mining identifies interpretable funding patterns (strongest rule lift = 6.30). The study contributes a reproducible example of how mature machine learning techniques can be systematically integrated to generate domain-specific evidence for research funding governance.

Keywords: machine learning; scientific projects; decision support; research funding; project analytics; data-driven management

1. Introduction

The increasing scale and complexity of scientific research projects, combined with growing competition for public funding, have significantly intensified the need for effective decision support in research governance. Modern research programmes typically involve international consortia, diverse organizational structures, and substantial financial resources, making the evaluation and management of project proposals a complex and

resource-intensive process [1–3]. Traditional assessment mechanisms, primarily based on expert review and qualitative judgment, remain central to funding decisions but face limitations in terms of scalability, consistency, and transparency when applied to large and heterogeneous project portfolios.

In recent years, the availability of large administrative and open datasets has created new opportunities for the application of data-driven methods in research management [4]. Machine learning techniques, in particular, have demonstrated considerable potential for extracting patterns from complex datasets and supporting evidence-based decision-making across various public-sector domains. Within the context of research funding, such approaches can complement expert evaluation by providing objective, reproducible insights derived from historical project data.

Despite these developments, existing analytical approaches to research funding remain largely fragmented. Prior studies have typically focused on isolated tasks, such as research performance evaluation, collaboration network analysis, or cost estimation, without offering an integrated project-level perspective on how project structure, financial configuration, and funding outcomes interact [5]. The present study therefore does not claim algorithmic novelty; instead, it addresses an application gap by combining mature analytical techniques into a systematic empirical workflow for research management. Project-level analyses are also constrained by data limitations, as information on rejected proposals is rarely available, complicating the formulation and interpretation of supervised learning tasks.

This study addresses these challenges by developing an integrated analytical pipeline that unifies three complementary tasks within a decision-support context: (1) identification of latent project structures via unsupervised learning; (2) prediction of project budgets via regression; and (3) classification of funding completeness via supervised learning. The contribution of the work lies in the systematic integration and interpretation of these tasks for scientific research management, rather than in introducing new algorithmic procedures. The three tasks are connected through a shared feature representation, and cluster membership from Stage 1 serves as an enriched input to Stage 3, forming a sequential pipeline that approximates a probabilistic model of funding outcomes under uncertainty.

The main objective of this paper is to assess whether structural and financial parameters of scientific projects contain sufficient information to support reliable, data-driven decision-making in research funding contexts. The central hypothesis is that project characteristics related to consortium composition, international collaboration, and budget configuration exhibit stable patterns that can be identified through unsupervised learning and can be used to predict budget scale and funding completeness with high accuracy.

Accordingly, the paper is positioned as an application-oriented study of methodological integration. To the best of the authors' knowledge, few project-level studies have systematically combined structure identification, budget prediction, funding-completeness classification, and association rule mining for scientific research management using participant-level and project-level CORDIS data.

To investigate this hypothesis, the study integrates three complementary analytical approaches. First, unsupervised learning methods are applied to identify latent project structures and typologies. Second, regression models are used to analyze the relationship between organizational features and total project budgets. Third, supervised classification and associative rule mining are employed to assess funding outcomes and extract interpretable patterns linked to successful and partially successful projects.

The results demonstrate that scientific projects form a limited number of stable structural configurations and that financial and organizational parameters play an important role in both budget determination and funding completeness. By integrating standard machine learning paradigms into a single empirical workflow, this study provides application-

https://doi.org/10.3390/app16147056

oriented evidence on the role of project structure in shaping funding outcomes and offers practical insights for the design of intelligent decision-support tools in research governance.

2. Related Work

The application of data-driven methods in research evaluation and science policy has expanded significantly over the past decade, driven by the growing availability of large-scale administrative and bibliometric datasets and the increasing complexity of research systems. Traditional approaches to assessing scientific activity have historically relied on expert judgment, peer review procedures, and descriptive statistical indicators. While these mechanisms remain central to research governance and funding allocation, their effectiveness is increasingly challenged in environments characterized by large project portfolios, international collaboration, and multidimensional performance criteria [6,7]. In particular, qualitative evaluation processes often struggle to scale efficiently and to provide consistent, reproducible insights when applied to heterogeneous sets of research projects with diverse organizational and financial configurations.

In response to these challenges, computational and analytical methods have gained prominence as complementary tools for supporting research management and funding decisions. Within the broader field of scientometrics and research analytics, statistical modeling and machine learning techniques have been widely adopted to analyze publication output, citation patterns, and collaboration dynamics [8–11]. These approaches have enabled the identification of structural regularities in scientific production and have contributed to evidence-based policy discussions regarding research performance, excellence, and impact. Network-based analyses, in particular, have emphasized the role of collaboration structures, co-authorship networks, and international cooperation in shaping scientific productivity and knowledge diffusion [12]. Such studies consistently demonstrate that structural characteristics of research systems influence outcomes, suggesting that similar analytical perspectives may be valuable at the level of research projects [13,14].

Despite these advances, much of the existing literature remains focused on output-oriented indicators, such as publications, citations, or journal metrics, rather than on the organizational and financial properties of scientific projects themselves [15]. Project-level analysis presents distinct methodological challenges, as projects represent temporary, goal-oriented organizational forms that combine financial resources, institutional capacities, and collaborative arrangements. Several strands of research have addressed specific analytical tasks at this level. Cost estimation and budget forecasting models have been developed to support project planning, often drawing on methods from engineering management and operations research. Separately, predictive models have been proposed to estimate research success or impact using proxy measures, such as publication volume or citation counts, as dependent variables [16]. While these studies provide valuable insights, they typically address isolated problems and do not offer an integrated analytical perspective that spans exploration, prediction, and evaluation.

A further limitation concerns the treatment of anomalous projects: many empirical studies exclude projects with exceptionally large budgets, risking biased conclusions regarding strategically important large-scale initiatives [17]. However, such anomalous projects often correspond to strategically important initiatives, including large collaborative infrastructures or mission-oriented research programs, which are of particular interest to funding agencies and policymakers [17]. The systematic exclusion of these cases risks producing biased conclusions and obscuring important governance challenges associated with large-scale scientific investments.

Recent contributions in the field of data-driven public administration and digital governance have argued for the development of intelligent decision-support systems

capable of integrating heterogeneous data sources and analytical methods to enhance transparency, efficiency, and accountability in public-sector decision-making [18]. Within this perspective, machine learning is increasingly viewed not merely as a predictive tool, but as a means of generating structured, interpretable insights that can inform strategic and operational decisions. Nevertheless, there remains a lack of comprehensive methodological frameworks combining unsupervised exploration, predictive modeling, and evaluative analysis in a coherent manner.

This study builds on prior research by addressing these limitations through an integrated machine learning-based analysis of scientific projects. The approach combines unsupervised learning for structure identification and anomaly detection, regression models for budget prediction, classification models for funding-completeness assessment, and association rule mining for interpretable pattern extraction. All technical tools used in the study are mature and well-established; accordingly, the contribution is positioned as methodological integration and domain-specific empirical insight for research governance rather than as the proposal of a new machine learning algorithm.

3. Data and Methods

The empirical basis of this study is formed by open data on scientific projects funded under the European Union's Sixth Framework Programme (FP6). The data were obtained from the Community Research and Development Information Service (CORDIS), which provides detailed administrative records on EU-funded research projects, including financial, organizational, and temporal characteristics. FP6 was selected for analysis due to the completeness and consistency of its project-level information, as well as its relevance as a large-scale international research funding programme.

The initial dataset comprised information on 8536 projects and 74,524 participating organizations. The data are organized into two interrelated tables: a project-level table containing general project attributes and a participant-level table describing organizations involved in each project. These tables were linked using a unique project identifier, enabling the aggregation of organizational information at the project level.

Following data cleaning and preprocessing procedures, the final analytical dataset consisted of 5278 projects launched between 2002 and 2008, involving organizations from 50 countries. The reduction in dataset size was primarily due to the removal of projects with incomplete or inconsistent financial information, which is essential for reliable budget modeling.

The data are organized into two interrelated tables linked by a unique project identifier (projectId). Table 1 describes the project-level table; Table 2 describes the participant-level table. Both tables were joined to aggregate organizational information at the project level.

Preprocessing was performed in five sequential steps. The rationale and parameters for each step are described below.

Step 1—Temporal encoding: Project start dates were converted into numerical year values (2002–2008).

Step 2—Financial variable standardization: Budget values were converted to uniform floating-point format; non-numeric characters removed via regular-expression cleaning.

Step 3—Organizational feature aggregation: Participant-level table aggregated at project level yielding: (1) num_organizations, (2) num_countries, (3) coordinator_country (proxy: most frequent participant country if no explicit coordinator).

Step 4—Missing value handling: Projects with missing/null financial data or missing coordinator country after proxy assignment were removed, reducing the dataset from 8536 to 5278 projects.

Step 5—Outlier removal: Projects above the 99th budget percentile were excluded (53 projects, ~1%). Sensitivity to 95th and 97th percentile thresholds produced comparable cluster structures. Final analytical dataset: 5225 projects.

The complete feature set used across all three analytical tasks is summarized in Table 3, which maps each feature to its notation in subsequent mathematical formulations.

Table 1. Data from CORDIS database.

Column	Data Type	Description
projectId	Integer	Unique project identifier (primary key)
acronym	String	Project acronym
title	String	Full title of the project
startDate	Date	Project start date
endDate	Date	Project end date
totalCost	Float	Total project budget (EUR)
ecMaxContribution	Float	Maximum EU financial contribution (EUR)
fundingScheme	String	Type of funding instrument
frameworkProgramme	String	Framework programme identifier (FP6)
topics	String	Thematic research topics/domains
status	String	Project status (CLOSED, TERMINATED, etc.)

Table 2. Participant-level table—attributes describing organizations involved in each project.

Column	Data Type	Description
projectId	Integer	Foreign key linking to project-level table
organizationId	Integer	Unique organization identifier
name	String	Name of the participating organization
country	String	Country of the organization (ISO code)
city	String	City of the organization
role	String	Role in project: Coordinator or Participant
ecContribution	Float	EU financial contribution to this organization (EUR)
activityType	String	Type of organization (HES, REC, PRC, PUB, OTH)
endOfParticipation	Date	Date organization left the project (if applicable)

Table 3. Complete feature set used for modeling (n = 5225 projects).

Feature Name	Symbol	Type	Used in	Description
num_organizations	x_1	Integer	Clustering, Regression, Classification	Number of participating organizations
num_countries	x_2	Integer	Clustering, Regression, Classification	Number of distinct countries
coordinator_country	x_3	Categorical	Clustering, Classification	Country of the lead organization
start_year	x_4	Integer	Regression, Classification	Project start year (2002–2008)
totalCost	x_5/y	Float (EUR)	Regression (target), Classification	Total project budget
ecMaxContribution	x_6	Float (EUR)	Regression, Classification	Maximum EU contribution
budget_per_org	$x_7 = x_5/x_1$	Float (EUR)	Classification	Budget per participating organization
internationalization	$x_8 = x_2/x_1$	Float	Clustering, Classification	Countries-to-organizations ratio
cluster_label	$c \in \{0, 1\}$	Binary	Classification (enriched feature)	Cluster assignment from unsupervised stage
funding_ratio	R = ec/total	Float [0, 1]	Target construction	Ratio of granted to requested funding
funding_outcome	$z \in \{0, 1\}$	Binary	Classification (target)	1 if $r \geq 0.70$, else 0

The three analytical tasks are connected through a sequential analytical pipeline designed to support empirical decision-making under uncertainty. For transparency, the workflow can be represented as an approximation of the posterior probability of funding outcome z_i given observable project characteristics x_i:

$$P(z_i \mid x_i) = \int P(z_i \mid x_i, c_i, y_i) \cdot P(c_i \mid x_i) \cdot P(y_i \mid x_i)\, dc_i\, dy_i \quad (1)$$

https://doi.org/10.3390/app16147056

where $c_i \in \{0, 1\}$ is the latent cluster assignment and y_i is the predicted budget. In operational terms, the workflow is factorized into three tractable steps: (1) unsupervised estimation of $P(c_i \mid x_i)$; (2) regression estimation of $E[y_i \mid x_i]$; and (3) supervised classification of $P(z_i \mid x_i, c_i, \hat{y}_i)$ using augmented features $[x_i, c_i, \hat{y}_i]$. This formulation is used to organize the empirical analysis and should be interpreted as an integration of existing methods, not as a new probabilistic algorithm.

The first stage identifies latent structural configurations among projects. Input features are $\{x_1, x_2, x_3, x_5, x_7, x_8\}$—organizational scale, geographic scope, financial parameters, and internationalization ratio. Dimensionality reduction via PCA, t-SNE, and UMAP is applied for visualization.

Hierarchical agglomerative clustering (HAC) with Ward's linkage minimizes total within-cluster variance at each merge step:

$$d(C_\alpha, C_\beta) = (n_\alpha \cdot n_\beta)/(n_\alpha + n_\beta) \cdot \|\mu_\alpha - \mu_\beta\|^2 \quad (2)$$

where C_α and C_β are clusters, n_α and n_β are their sizes, and μ_α, μ_β are centroids. The cluster label $c_i \in \{0, 1\}$ is assigned to each project and passed to the classification stage as an enriched feature. Anomaly detection flags projects whose distance to the nearest centroid exceeds the 99th percentile.

Figure 1 illustrates the relationship between predicted and actual project budgets for the linear regression model. The distribution of points around the ideal prediction line demonstrates that the model provides accurate estimates for small and medium-scale projects, which constitute the majority of the dataset. For larger projects, the dispersion of prediction errors increases, indicating greater uncertainty and reduced model precision at higher budget levels. This pattern reflects the presence of unobserved factors influencing large-scale projects, such as thematic priorities or strategic policy considerations, which are not captured by the available features.

Feature importance analysis further highlights the dominant role of consortium size in budget prediction. The number of participating organizations accounts for more than 60% of the explanatory power in the ensemble models, followed by measures of international collaboration intensity. The country of the coordinating organization and the project start year contribute comparatively less to budget variation.

Building on the descriptive statistics and regression-based budget analysis presented above, further insights into the heterogeneity of FP6 projects were obtained through unsupervised learning techniques. These methods were employed to examine whether the observed variation in budgets and consortium characteristics corresponds to stable structural patterns within the dataset, rather than representing purely continuous variability.

The analysis confirms that FP6 projects are not distributed uniformly in the feature space. Instead, distinct structural configurations emerge when organizational scale, international composition, and financial parameters are considered jointly. Dimensionality reduction results indicate that a substantial share of variance can be explained by a limited number of latent dimensions, while nonlinear projections reveal clearer separation between groups of projects with different structural profiles. This suggests that project characteristics cluster around a small number of recurring configurations rather than forming a single homogeneous population. Figure 2 provides a comparative visualization of linear and nonlinear dimensionality reduction techniques applied to the FP6 project dataset.

https://doi.org/10.3390/app16147056

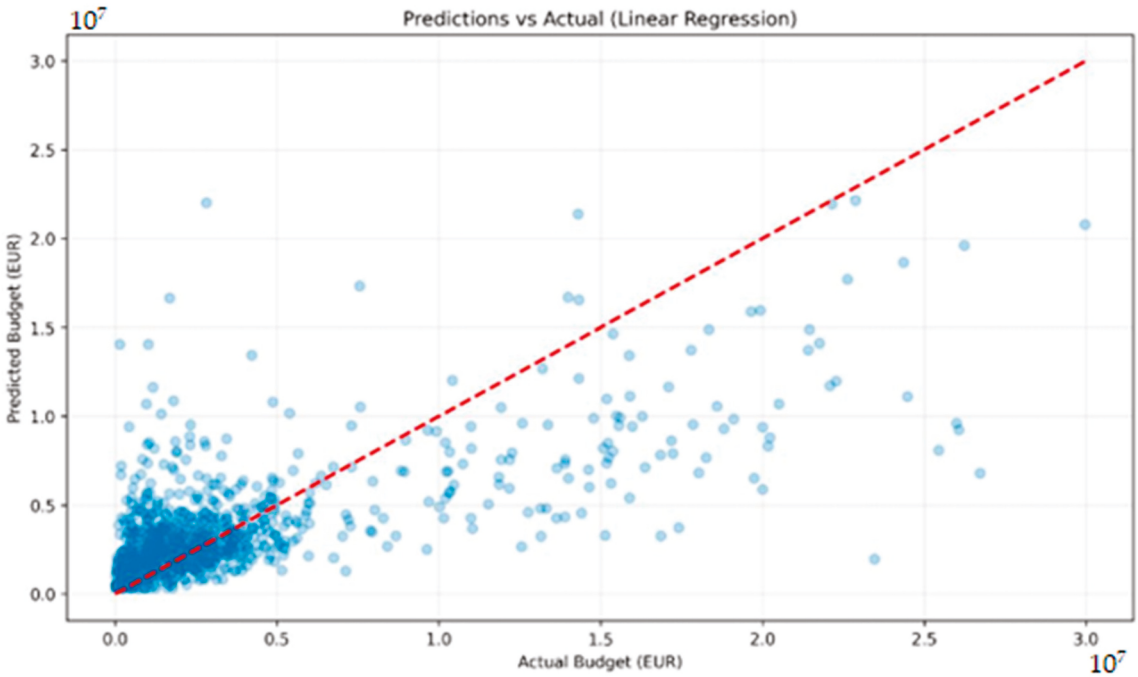


Figure 1. Predicted vs. actual project budgets using the linear regression model.

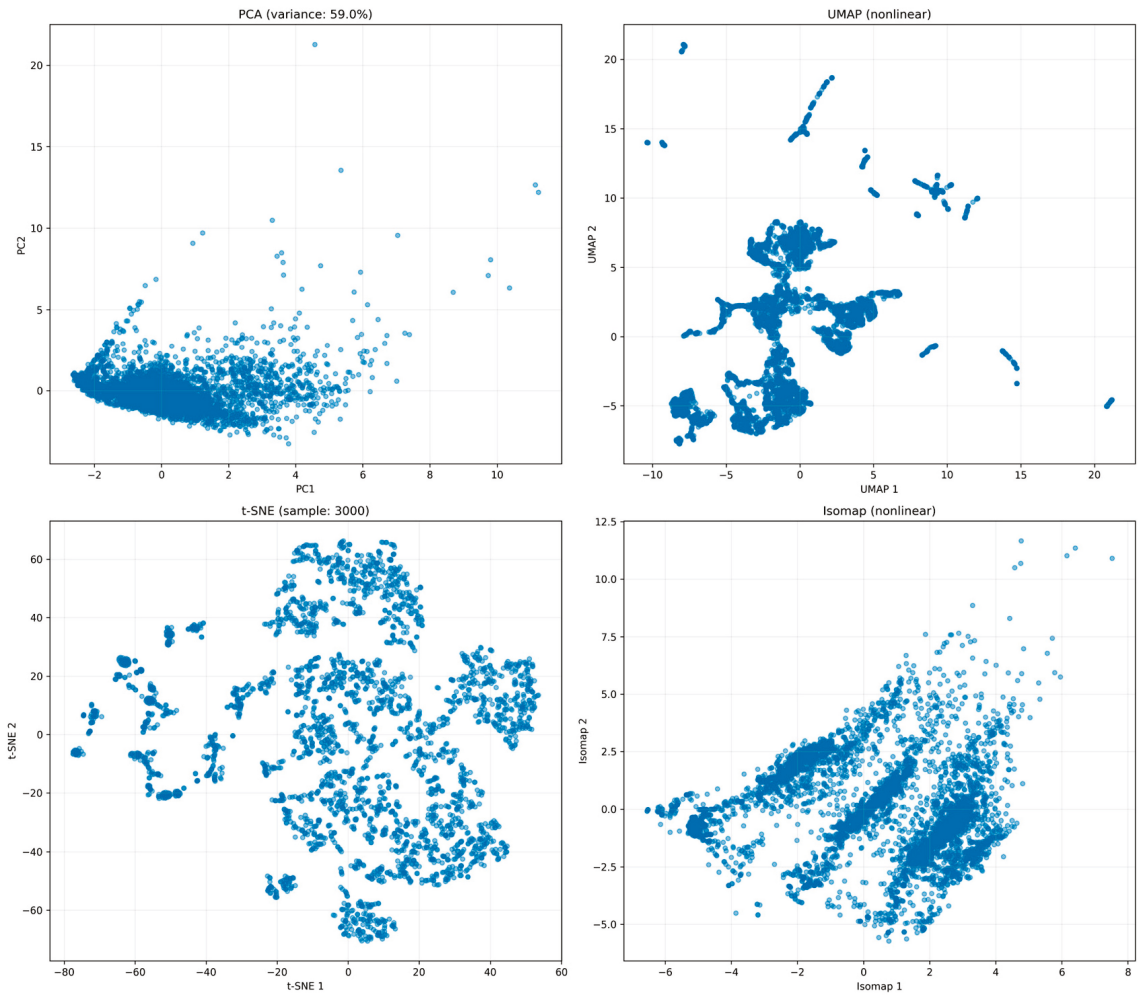


Figure 2. Dimensionality reduction comparison (FP6 dataset).

The comparison of dimensionality reduction techniques illustrates that while linear projection captures dominant variance directions, nonlinear methods provide a more informative representation of project heterogeneity. UMAP and t-SNE reveal compact groups and isolated observations, supporting the existence of distinct project configurations rather than a continuous distribution of characteristics.

To further examine whether the structural patterns revealed by dimensionality reduction correspond to stable and interpretable groupings of projects, hierarchical clustering was applied to the reduced feature space. While dimensionality reduction techniques demonstrated that FP6 projects occupy a heterogeneous and non-uniform space, clustering analysis enables the identification of discrete project groupings characterized by similar organizational and financial configurations.

The second stage estimates total project cost y_i from features $\{x_1, x_2, x_3, x_4, x_6\}$:

$$y_i = f(x_i) + \varepsilon_i \tag{3}$$

where y_i is the total budget (EUR), $x_i \in \mathbb{R}^m$ (m = 5) is the input feature vector, and ε_i is a stochastic error term capturing unobserved factors.

For the linear baseline:

$$y_i = \beta_0 + \Sigma_{j=1}{}^{m} \beta_j x_{ij} + \varepsilon_i \tag{4}$$

Coefficients β_j are estimated by minimizing the residual sum of squares:

$$\min_\beta \ \Sigma_{i=1}{}^{n} (y_i - \hat{y}_i)^2 \tag{5}$$

Ensemble models approximate f as a weighted combination of K decision trees $h_k(x)$:

$$f(x) = \Sigma_{k=1}{}^{K} w_k h_k (x) \tag{6}$$

Overfitting is diagnosed when training error is substantially lower than test error:

$$E_train[(y_i - \hat{y}_i)^2] \ll E_test[(y_i - \hat{y}_i)^2] \tag{7}$$

Table 4 presents the main characteristics of the two identified clusters. The first cluster comprises approximately 72% of projects and is characterized by relatively small budgets, compact consortia, and higher internationalization ratios. The second cluster, accounting for roughly 28% of projects, corresponds to large-scale initiatives with significantly higher budgets, substantially larger consortia, and a lower relative level of internationalization per participating organization. The magnitude of the differences across clusters indicates that the identified grouping captures meaningful distinctions in project scale and organizational structure rather than marginal variation.

Table 4. Cluster results.

Characteristic	Cluster 0 (Small Projects)	Cluster 1 (Large Projects)	Relationship
Number of projects	~3700 (72%)	~1400 (28%)	2.6:1
Average budget	1.62 M EUR	10.41 M EUR	6.4×
Median budget	1.25 M EUR	8.74 M EUR	7.0×
Average number of organizations	6.75	21.44	3.2×
The median of organizations	7	20	2.9×
Average number of countries	4.48	10.13	2.3×
Median of countries	5	10	2.0×
The level of internationalization	0.75	0.49	0.7×
The budget for the organization	260 K EUR	539 K EUR	2.1×

https://doi.org/10.3390/app16147056

The target variable is constructed from the funding ratio r_i = ecMaxContribution$_i$/totalCost$_i$:

$$z_i = 1 \text{ if } r_i \geq 0.70, \text{ else } 0 \tag{8}$$

The threshold $\tau = 0.70$ was selected based on the empirical distribution of r_i in the FP6 dataset, where approximately 65% of projects receive coverage ≥ 0.70, creating a meaningful and non-degenerate binary split. Alternative values $\tau \in \{0.60, 0.65, 0.75, 0.80\}$ were also evaluated; all produced qualitatively similar results with accuracy differences < 2 percentage points. Since CORDIS contains only approved projects, z = 0 indicates a budget adjustment rather than project rejection—explicitly acknowledged as a limitation.

The classification feature vector is the augmented set $[x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, c]$. Six algorithms are evaluated: logistic regression, decision tree, random forest, gradient boosting, XGBoost, MLP, and k-NN. All models are trained using stratified 5-fold cross-validation.

Model performance is assessed using accuracy, F1-score, and AUC:

$$\text{AUC} = \int_0^1 \text{TPR}(t)\, d(\text{FPR}(t)) \tag{9}$$

where higher AUC values indicate better classification performance.

Figure 3 illustrates the ROC curves for all evaluated models. The results demonstrate consistently high predictive performance, with most models achieving AUC values above 0.97, indicating strong separability between fully and partially funded projects.

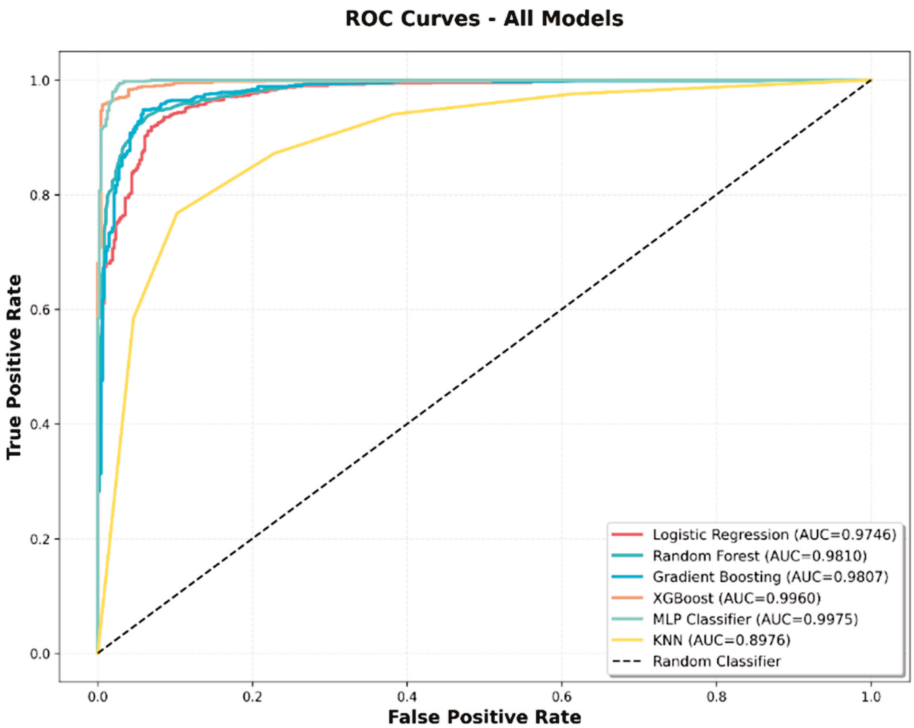


Figure 3. ROC curves for all evaluated classification models, demonstrating high predictive performance and strong separability between fully and partially funded projects.

Across all evaluated models, classification results demonstrate consistently high predictive performance, indicating that funding-completeness categories are strongly associated with observable project characteristics in the FP6 dataset. All six models achieved accuracy values above 83%, confirming systematic relationships between consortium structure, financial parameters, and the constructed funding-completeness target. The strongest

performance was obtained using nonlinear models, particularly the multilayer perceptron and gradient boosting-based approaches; however, these results are interpreted as evidence of structural regularity in the data rather than as proof of a new methodological advance.

Association rule mining is applied to the discretized feature space. Features x_1, x_2, x_5, x_6 are binned into low/medium/high categories (tertile thresholds). The Apriori algorithm is run with minimum support = 0.05 and minimum confidence = 0.60. Rules are ranked by lift:

$$\text{lift}(A \rightarrow B) = P(A \cup B)/(P(A) \cdot P(B)) \tag{10}$$

The multilayer perceptron achieved the highest overall performance, with an accuracy of 98.50%, an F1-score of 98.39%, and a ROC-AUC of 0.9975. This model exhibited a very low error rate, with only two false negatives in the test set, indicating a high level of sensitivity in identifying fully funded projects. XGBoost demonstrated similarly strong performance, achieving an accuracy of 96.84% and the most stable cross-validation results, suggesting robustness to variations in the training data. In contrast, linear logistic regression achieved lower but still substantial accuracy (90.92%), indicating that a significant portion of the decision boundary is linearly separable, while nonlinear interactions further improve classification performance.

Analysis of ROC curves shows that all models, except k-nearest neighbors, exhibit strong discriminatory power, with ROC-AUC values exceeding 0.97. The limited performance of k-nearest neighbors reflects the heterogeneous and unevenly distributed nature of projects in the feature space, which was previously observed in the unsupervised analysis. Importantly, misclassification errors are concentrated among projects with intermediate budgets and consortium sizes, while projects at the extremes of the structural spectrum are classified with substantially higher accuracy.

To complement predictive results with interpretable insights, associative rule mining was applied to identify frequent patterns linked to funding outcomes. Rules with the highest lift values reveal that projects involving many participating countries are substantially more likely to receive full funding, while excessively large consortia combined with high requested budgets are strongly associated with partial funding outcomes. The strongest rule for partial funding exhibits a lift of 6.30, indicating a markedly increased likelihood of reduced funding for overly ambitious project configurations.

Finally, feature importance analysis across multiple models confirms the dominance of financial variables in funding-completeness prediction. Total project cost, maximum EU contribution, and budget per organization jointly account for approximately 69% of explanatory importance, while consortium size and thematic domain provide additional discriminatory power. The convergence of results across classification models, association rules, and feature importance measures reinforces the empirical reliability of the identified patterns and highlights the central role of structural and financial configuration in determining funding completeness.

4. Results

Dimensionality reduction confirms that FP6 projects are not uniformly distributed in the feature space. PCA reveals that the first two principal components explain approximately 68% of total variance. Figure 1 shows the PCA projection, while Figure 2 presents the UMAP projection. The cluster profile comparison of key characteristics is presented in Table 5.

Table 5. Cluster profile comparison of key characteristics.

Characteristic	Cluster 0 (Small Projects)	Cluster 1 (Large Projects)	Ratio (C1/C0)
Number of projects	~3700 (72%)	~1525 (28%)	—
Mean totalCost (x_5)	1.62 M EUR	10.41 M EUR	6.4×
Median totalCost (x_5)	1.25 M EUR	8.74 M EUR	7.0×
Mean num_organizations (x_1)	6.75	21.44	3.2×
Mean num_countries (x_2)	4.48	10.13	2.3×
Mean internationalization (x_8)	0.75	0.49	0.65×
Mean budget_per_org (x_7)	260 K EUR	539 K EUR	2.1×

Cluster 0 (72%): compact consortia, higher internationalization, modest budgets. Cluster 1 (28%): large-scale initiatives with substantially larger consortia and budgets. Differences are statistically significant (Mann–Whitney U, $p < 0.001$). Anomaly detection identified 41 atypical projects; none were misclassified in the supervised stage.

Table 6 summarizes regression model performance on the held-out test set (n = 1045 projects). Confidence intervals computed via bootstrap resampling (1000 iterations).

Table 6. Regression model performance on the test set.

Model	R^2 (Test)	95% CI	RMSE (EUR)	Train R^2	Overfitting?
Linear Regression	0.499	[0.471, 0.527]	2,140,000	0.502	No
Ridge Regression (λ = 1.0)	0.501	[0.473, 0.529]	2,135,000	0.503	No
Random Forest (K = 100)	0.503	[0.475, 0.531]	2,128,000	0.741	Yes
Gradient Boosting	0.488	[0.461, 0.516]	2,167,000	0.782	Yes

All models achieve $R^2 \approx 0.50$ on the test set. This result warrants critical interpretation: the unexplained variance (≈50%) reflects latent factors not captured in administrative data—thematic priorities, geopolitical considerations, and qualitative aspects of the research agenda. Linear and ridge models generalize well (train ≈ test R^2), confirming an approximately linear budget–organization relationship. Ensemble methods overfit without improving test performance.

All models achieve accuracy above 83%, with nonlinear models reaching 96–98%. The high AUC values (>0.97) raise a legitimate concern about potential data leakage and dependence between predictors and the constructed target variable. Models retrained excluding x_6 (ecMaxContribution) yielded AUC = 0.91–0.95, which substantially mitigates, but does not fully eliminate, this concern. The remaining high performance is therefore interpreted cautiously as evidence of structural regularity within the FP6 programme rather than as a universally generalizable result. Cross-validation results confirm stability across folds (CV Std < 0.01 for best models).

Association rule mining identified 127 rules meeting minimum thresholds. Table 7 presents the eight highest-lift rules; the full rule set.

Table 7. Top-eight association rules by lift value.

Antecedent	Consequent	Support	Confidence	Lift
high totalCost ∧ high num_org	z = 0 (partial)	0.08	0.81	6.30
high num_countries ∧ low budget_per_org	z = 1 (full)	0.12	0.87	4.92
cluster = 1 ∧ high totalCost	z = 0 (partial)	0.10	0.79	4.51
low num_org ∧ high internationalization	z = 1 (full)	0.14	0.83	4.38
medium totalCost ∧ medium num_org	z = 1 (full)	0.18	0.78	3.97
high ecMaxContribution ∧ cluster = 1	z = 0 (partial)	0.07	0.75	3.84
low totalCost ∧ high num_countries	z = 1 (full)	0.09	0.80	3.72
high num_org ∧ low internationalization	z = 0 (partial)	0.06	0.72	3.61

https://doi.org/10.3390/app16147056

5. Discussion and Limitations

The central contribution of the study is not the introduction of new machine learning algorithms, but the systematic integration of mature analytical tools for project-level research funding analysis. The stable two-cluster structure separating small-scale and large-scale projects is robust across the analytical stages. The dominance of consortium size (x_1) as the primary budget predictor (61% of feature importance) indicates that financial requirements in FP6 are driven by coordination complexity, consistent with prior research on collaboration structures in science [12,13].

The $R^2 \approx 0.50$ for budget regression deserves explicit discussion. The unexplained half of budget variance likely reflects thematic and strategic factors absent from administrative records: research priority areas, geopolitical considerations, and qualitative project characteristics that influence funding decisions but leave no trace in the CORDIS data. This finding itself carries a governance implication—richer data sources (topic models applied to project abstracts, reviewer scores) could substantially improve explanatory power.

The high AUC values (>0.97) are unusual and require careful interpretation. As discussed above, removing ecMaxContribution reduces AUC to 0.91–0.95, indicating partial but not complete dependence on a target-correlated feature. The residual high performance appears to reflect the structural regularity of the FP6 programme, but this result may not generalize to programmes with different funding logics. External validation is therefore essential before applying the workflow as an operational decision-support tool.

First, the analysis is based on a single funding programme (FP6); findings may not transfer directly to FP7, Horizon 2020, or Horizon Europe. External validation was not performed and constitutes a priority for future work.

Second, the operationalization of funding success as $r \geq 0.70$ conflates budget adjustment with funding completeness. Within a dataset of only approved projects, $z = 0$ indicates a negotiated budget reduction, not project failure.

Third, the feature set is restricted to administrative and organizational descriptors. Thematic content, scientific quality signals, and applicant track records are unavailable in CORDIS open data but are known to influence outcomes.

Fourth, comparison with non-ML baseline methods was not performed. Such comparison would be necessary to empirically demonstrate the added value of the integrated analytical pipeline relative to descriptive statistics or expert-only assessment procedures.

Overall, the primary contribution lies in methodological integration and empirical insights: the study shows how structure identification, budget prediction, funding-completeness classification, and association rule mining can be combined within one reproducible workflow for research management. Future work should pursue: (1) external validation on FP7/Horizon 2020; (2) enrichment with topic models; (3) comparison with expert evaluation and non-ML baselines; (4) development of a fully Bayesian probabilistic formulation; and (5) fairness metric analysis for geographic and organizational biases.

6. Conclusions

This paper presented an integrated machine learning-based analysis applied to 5225 projects from the CORDIS FP6 database. The study does not claim to develop new algorithms; instead, it demonstrates how established methods can be combined to generate empirical insights for scientific project governance. Three key conclusions emerge.

First, FP6 projects form two structurally distinct configurations. Consortium size is the dominant driver of both budget scale and cluster membership, accounting for over 60% of predictive importance in regression models.

Second, budget prediction is achievable with $R^2 \approx 0.50$ using organizational and collaboration features alone. The 50% unexplained variance reflects genuine latent fac-

https://doi.org/10.3390/app16147056

tors absent from administrative data, not a modeling failure. Ensemble methods do not outperform linear models on the test set.

Third, funding completeness is predictable with high accuracy (AUC > 0.97 for most models), but this result should be interpreted cautiously because the target variable is constructed from financial variables available in the dataset. Feature ablation experiments reduce, but do not entirely remove, this concern. Association rule mining identifies interpretable patterns: high country diversity with low budget-per-organization predicts full funding; large consortia with high total costs are associated with partial funding.

These findings demonstrate that integrated machine learning analysis can support evidence-based decision-making in research funding governance when used as a complementary tool alongside expert assessment. The value of the work lies in the systematic integration of analytical tasks and the domain-specific insights obtained from FP6 data, subject to the limitations discussed and the external validation that future work should provide.

Author Contributions: Conceptualization, A.M. (Arailym Medetbek), A.M. (Aidos Mukhatayev) and A.B.; methodology, A.M. (Arailym Medetbek) and S.B.; software, D.B.; validation, A.B., S.B. and A.M. (Aidos Mukhatayev); formal analysis, A.B.; investigation, A.M. (Arailym Medetbek); resources, A.M. (Arailym Medetbek); data curation, D.B.; writing—original draft preparation, A.M. (Arailym Medetbek); writing—review and editing, A.M. (Aidos Mukhatayev) and S.B.; visualization, D.B.; supervision, A.B. and A.M. (Aidos Mukhatayev). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in European data at https://data.europa.eu/data/datasets/cordisfp6projects?locale=en (accessed on 4 April 2026).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. European Union Open Data Portal. CORDIS FP6 Projects Dataset. Available online: https://data.europa.eu/data/datasets/cordisfp6projects (accessed on 4 April 2026).
2. European Commission. Sixth Framework Programme (FP6), 2002–2006. Available online: https://cordis.europa.eu/programme/id/FP6 (accessed on 4 April 2026).
3. Biloshchytskyi, A.; Kuchansky, A.; Andrashko, Y.; Biloshchytska, S.; Kuzka, O.; Shabala, Y.; Lyashchenko, T. A method for the identification of scientists' research areas based on a cluster analysis of scientific publications. *East.-Eur. J. Enterp. Technol.* **2017**, *5*, 4–11. [CrossRef]
4. Bornmann, L. Measuring impact in research evaluations: A thorough discussion of methods for, effects of, and problems with impact measurements. *High Educ.* **2017**, *73*, 775–787. [CrossRef]
5. Zhang, T.; Qin, S. Structural Deformation Prediction and Uncertainty Quantification via Physics-Informed Data-Driven Learning. *Appl. Sci.* **2026**, *16*, 3194. [CrossRef]
6. Biloshchytskyi, A.; Omirbayev, S.; Mukhatayev, A.; Kuchanskyi, O.; Biloshchytska, S.; Andrashko, Y.; Toxanov, S.; Faizullin, A. A structural model for building a system for the development of methodological competence and methods for evaluating its effectiveness. *East.-Eur. J. Enterp. Technol.* **2023**, *5*, 6–22. [CrossRef]
7. Reis, V.B.; Baptista, J.S.; Duarte, J. Machine Learning in Surface Mining—A Systematic Review. *Appl. Sci.* **2026**, *16*, 3246. [CrossRef]
8. Biloshchytskyi, A.; Omirbayev, S.; Mukhatayev, A.; Faizullin, A.; Toxanov, S.; Kassenov, K. Research on the formation level of methodological competence of it disciplines teachers. In *Proceedings of the 2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine, 25–27 November 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 242–245. [CrossRef]
9. Liu, Z.; Fan, L.; Chen, Q.; Li, H.; Wei, A. Performance Evaluation of Publicly Funded Agricultural Research Projects with Light-TabNet. *Appl. Sci.* **2026**, *16*, 3230. [CrossRef]

10. Liu, J.; Gratz, J.; Amour, C.; Nshama, R.; Walongo, T.; Maro, A.; Mduma, E.; Platts-Mills, J.; Boisen, N.; Nataro, J.; et al. Optimization of Quantitative PCR Methods for Enteropathogen Detection. *PLoS ONE* **2016**, *11*, e0158199. [CrossRef] [PubMed]
11. Biloshchytskyi, A.; Kuchanskyi, O.; Mukhatayev, A.; Andrashko, Y.; Toxanov, S.; Faizullin, A.; Kassenov, K. Application of Time-Weighted PageRank Method with Citation Intensity for Assessing the Recent Publication Productivity and Partners Selection in R&D Collaboration. *Publications* **2024**, *12*, 48. [CrossRef]
12. Bertsimas, D.; Kallus, N. From predictive to prescriptive analytics. *Manag. Sci.* **2020**, *66*, 1025–1044. [CrossRef]
13. Biloshchytskyi, A.; Omirbayev, S.; Mukhatayev, A.; Biloshchytska, S.; Toxanov, S.; Faizullin, A. The concept of the Internet of Things in the development of information and analytical systems based on the method of constructing a scalar assessment of the results of research activities of scientists. *Procedia Comput. Sci.* **2024**, *231*, 684–690. [CrossRef]
14. He, C.; Duan, Y.; Mamat, T.; Zhu, X.; Dridi, M.; Mualla, Y.; Abbas-Turki, A. Research on a Prediction Method for Maintenance Decision of Expressway Asphalt Pavement Based on Random Forest. *Appl. Sci.* **2026**, *16*, 3343. [CrossRef]
15. Ashraf, R.; Khan, M.A.; Khuhro, R.A.; Bhatti, Z.A. Knowledge creation dynamics of technological forecasting and social change special issues. *Technol. Forecast. Soc. Change* **2022**, *180*, 121663. [CrossRef]
16. OECD. Artificial Intelligence, Machine Learning and Big Data in Public Policy. 2021. Available online: https://www.oecd.org/gov/digital-government/ai-ml-big-data-public-policy.htm (accessed on 5 April 2026).
17. Okamura, K. Scientometric engineering: Exploring citation dynamics via arXiv eprints. *Quant. Sci. Stud.* **2022**, *3*, 122–146. [CrossRef]
18. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI. *Nat. Mach. Intell.* **2020**, *2*, 252–259. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

applied sciences

MDPI

Article

Integrated Production and Transportation Scheduling in a Distributed Hybrid Flow Shop Considering Dynamic Orders Under an E-Commerce Environment

Ziyang Jin and Meiyan Li *

School of Energy and Mining Engineering, Shandong University of Science and Technology, Qingdao 266590, China; 202483010184@sdust.edu.cn
* Correspondence: skd994322@sdust.edu.cn

Abstract

Real-time order arrivals and stringent timeliness demands in e-commerce pose significant challenges to production-distribution coordinated scheduling in distributed manufacturing systems. This paper tackles the integrated production and distribution scheduling problem within a dynamic distributed hybrid flow shop. We formulate a mixed-integer linear programming model aimed at minimizing average order tardiness and total operational cost. To address dynamic uncertainties, we propose an improved decomposition-based multi-objective evolutionary algorithm (I-MOEA/D) operating within a rolling horizon framework. The algorithm integrates three critical components: an urgency-based emergency window management strategy that caps computational complexity, a 2-opt local search operator that boosts vehicle routing efficiency, and acceleration techniques—including order dictionary indexing, lightweight object replication, and distance caching—that guarantee real-time responsiveness. The window size is determined as W_{max} = 50 through sensitivity analysis. Extensive experiments conducted on 100, 200, and 300-order scenarios with 30 independent random seeds demonstrate that I-MOEA/D markedly outperforms NSGA-II, MOALNS, and a reinforcement learning-driven hyper-heuristic (RL-HH). For 200 orders, I-MOEA/D reduces average tardiness by 18.5%, 41.0%, and 58.7% compared to NSGA-II, MOALNS, and RL-HH, respectively, while maintaining competitive cost performance (a cost gap of 16.99% relative to the static lower bound). The IGD metric achieves 0.0031, confirming good convergence and diversity. Acceleration techniques cut computation time by 56% without sacrificing solution quality. The algorithm scales efficiently with problem size. These results validate that I-MOEA/D delivers effective real-time decision support for dynamic distributed production-distribution systems.

Keywords: distributed scheduling; hybrid flow shop; production-distribution; coordinated scheduling; dynamic scheduling; multi-objective evolutionary algorithm based on decomposition (MOEA/D); rolling horizon

1. Introduction

Against the backdrop of e-commerce globalization and the emergence of the Industry 5.0 paradigm, modern manufacturing has undergone a fundamental shift from traditional mass production to high-frequency, mass-customization order fulfillment models [1]. This transformation compels enterprises not only to manufacture products efficiently but also to deliver finished goods to geographically dispersed end customers within stringent

time windows. As a consequence, the integration of multiple geographically distributed hybrid flow shops (DHFS) has become the dominant operational infrastructure of contemporary supply chains [2]. However, production scheduling at the shop-floor level and vehicle routing at the logistics level have traditionally been treated as mutually independent decision problems. This sequential optimization approach often gives rise to severe systemic bottlenecks—finished goods accumulating on shop floors awaiting transport vehicles, or trucks making empty trips due to upstream production delays [3]. Against this background, integrated production and transportation coordinated scheduling has emerged as a critical research frontier in supply chain management [4].

In recent years, the Integrated Production and Distribution Scheduling Problem (IPDSP) has attracted considerable research attention. Chen, in a seminal review, systematically examined integrated production and outbound distribution scheduling, highlighting that collaborative optimization can significantly reduce operational costs and improve customer satisfaction [5]. A recent extension by Chen further surveys online IPDS problems, underscoring the growing importance of real-time decision-making in dynamic environments [6]. Liu and Liu provided a systematic survey of assembly flow shop scheduling and its extensions to distributed manufacturing and integrated supply chain, covering foundational models, multi-plant scheduling, and coupling with upstream material supply and downstream distribution [7]. Early studies predominantly focused on single-factory settings. Mohammadi et al. investigated coordinated production and distribution scheduling in a flexible job shop environment within a furniture manufacturing enterprise, establishing a bi-objective mixed-integer programming model [8]. Luo et al. designed an improved NSGA-II algorithm for a three-stage production–inventory–distribution integrated scheduling problem in a homogeneous flow shop, simultaneously optimizing earliness/tardiness penalty costs and energy consumption [9]. Long et al. proposed a hierarchical multi-objective particle swarm optimization algorithm (LMPSO) for integrated scheduling problems incorporating inventory holding, distribution, and tardiness costs [10]. More recently, Horváth provided new computational results for short-lifespan products in a single-machine single-vehicle setting and extended the analysis to multi-plant configurations [11]. Sidki and Tchernev developed a monolithic batch-centric MILP approach for integrated production and pipeline distribution scheduling, validated on real-world phosphate supply chain instances [12]. Additionally, Kaviyani-Charati and Defersha addressed integrated multi-line flexible flowshop production and multi-trip vehicle routing with minimum life-on-receipt constraints, proposing a hybrid genetic algorithm with local search [13]. Azad and Rahman developed a bi-objective model and a hybrid NSGA-II for sustainable production scheduling and routing in supply chains, minimizing tardiness and CO_2 emissions [14]. Erdogan and Dang proposed a matheuristic for integrated scheduling of manufacturing assembly, product transport, and material supply, coordinated by AGVs and AMRs, validated on a battery manufacturing case study [15]. Wang and Zhang designed a knowledge-guided evolutionary algorithm for integrated group production and transportation scheduling with inter-cell coordination in cellular manufacturing systems [16].

With the advancement of globalized production, distributed manufacturing has attracted increasing scholarly interest. Duan et al. provided a comprehensive review of distributed shop scheduling, covering the classification, modeling, and algorithmic approaches for distributed parallel machine, flow shop, job shop, and assembly shop scheduling [17]. Fu et al. studied coordinated production and distribution scheduling in a multi-factory permutation flow shop environment, proposing an improved black widow optimization algorithm [18]. Cai et al. designed a hybrid shuffled frog-leaping algorithm combined with Q-learning for the distributed assembly hybrid flow shop scheduling problem [2]. Shao et al. proposed a constructive heuristic based on the min-median rule and a multi-

neighborhood iterative greedy algorithm for the distributed hybrid flow shop scheduling problem (DHFSP) [19]. Wang and Wang developed a bi-population cooperative memetic algorithm (BCMA) for DHFSP in heterogeneous factory environments, substantially outperforming Gurobi and existing iterative greedy algorithms [20]. Liu et al. considered sequence-dependent setup times in the distributed heterogeneous permutation flow shop scheduling problem (DHPFSP-SDST) and employed Q-learning to dynamically select local search operators [21]. Tibaldo et al. established an MILP model that simultaneously optimizes batch splitting, production, and distribution scheduling in single-stage batch plants [22]. Mzili et al. proposed a hybrid genetic and penguin search optimization algorithm (GA-PSeOA) for efficient flow shop scheduling, demonstrating superior performance over pure GA and other metaheuristics [23].

Despite these advances, the vast majority of existing research remains confined to static deterministic environments, assuming that all order information is fully known prior to scheduling commencement [19]. In real-world e-commerce operations, however, customer orders arrive dynamically and randomly over a continuous time horizon [24]. Conventional static scheduling methods quickly become ineffective in such scenarios.

To address this challenge, the rolling horizon strategy has been widely adopted to handle uncertainties such as random order arrivals. This strategy partitions the scheduling timeline into consecutive decision windows, solves a local static subproblem within each window, and achieves dynamic responsiveness through periodic rescheduling. Ouelhadj and Petrovic provided a comprehensive survey of dynamic scheduling in manufacturing systems, elaborating on heuristics, metaheuristics, multi-agent systems, and AI-based methods [24]. Ding et al. presented a state-of-the-art survey on AI-enabled dynamic manufacturing scheduling, covering both directed heuristics and autonomous learning methods [25]. Yang et al. proposed a concurrent optimization method for real-time scheduling and reconfiguration based on deep reinforcement learning (DRL), addressing dynamic job arrivals and machine breakdowns [26].

At the algorithmic level for multi-objective dynamic scheduling, MOEA/D and NSGA-II represent the two most widely adopted methods. NSGA-II achieves a balance between elitism and population diversity through fast non-dominated sorting and crowding distance [27]. MOEA/D decomposes a multi-objective problem into a set of scalar subproblems that are solved cooperatively, demonstrating excellent convergence and distribution uniformity [28]. Recent surveys have systematically reviewed the development of decomposition-based evolutionary multi-objective optimization (EMO) from its inception to the current state of the art [29]. Wang et al. proposed an improved MOEA/D for low-carbon many-objective flexible job shop scheduling, demonstrating its effectiveness in complex scheduling environments [30].

With respect to e-commerce order fulfillment, Chen et al. studied task assignment and scheduling in multi-threaded electronic order fulfillment, proposing a depth-first heuristic algorithm [31]. Ren et al., through a systematic review of 378 publications, constructed an omni-channel logistics decision-making framework encompassing 45 decision elements, observing that the closer the logistics segment is to the end consumer, the higher its dynamism and complexity [32]. Recent work has also explored reinforcement learning-based evolutionary algorithms for distributed assembly permutation flow shops, incorporating Q-Learning and Sarsa to select high-quality local search operators [33]. Nevertheless, research that integrates dynamic order arrivals into the coordinated production and distribution scheduling of distributed hybrid flow shops remains extremely scarce. This problem necessitates simultaneous consideration of: (1) dynamic assignment of orders among heterogeneous factories; (2) dynamic scheduling in hybrid flow shops with multiple stages; (3) dynamic planning of capacitated multi-depot vehicle routing; and (4) rolling reschedul-

ing with schedule stability assurance. The coupling of these four dimensions induces exponential growth in problem complexity, rendering traditional static methods inadequate.

In light of the foregoing literature review, research that incorporates dynamic and stochastic order arrivals into production–distribution coordinated scheduling in distributed hybrid flow shops remains conspicuously lacking. This gap has become particularly pronounced amid the rapid expansion of e-commerce. During peak promotional events such as "Double 11" (Singles' Day), orders surge into the system with high density and uncertainty within a short time window, forcing factories and logistics networks to make rapid scheduling decisions under severely incomplete information. Any static, offline scheduling scheme would become obsolete within minutes, resulting in widespread tardiness, escalating costs, and diminished customer satisfaction.

To bridge this gap, this paper formulates the problem as a bi-objective dynamic optimization problem under a rolling horizon framework. The main contributions are summarized as follows.

First, at the problem formulation level. We simultaneously account for the dynamic arrival characteristics of orders and the integrated optimization of production and distribution within a distributed hybrid flow shop environment. An MILP model is established with the objectives of minimizing average order tardiness and total operational cost, capturing the coupled "dynamic assignment, multi-stage processing, and multi-depot delivery" features inherent in e-commerce fulfillment.

Second, at the algorithmic level. We propose an improved decomposition-based multi-objective evolutionary algorithm (I-MOEA/D) embedded within a rolling horizon architecture. Core enhancements include: an emergency window management strategy based on due-date urgency to prioritize critical orders while controlling computational scale; a 2-opt local search operator for vehicle delivery routes to reduce transportation costs; and acceleration techniques—comprising order dictionary indexing, lightweight object replication, and distance caching—that reduce single-window optimization to the order of seconds, thereby meeting real-time rescheduling requirements.

Third, with respect to practical value. Once an order is scheduled, its factory assignment is immediately committed and never revised in subsequent windows. This immediate commitment policy ensures that system nervousness—defined as the proportion of orders whose assignment changes between consecutive windows—is zero by design. This stability is a deliberate trade-off: sacrificing a degree of adaptability to future arrivals in exchange for deterministic, disturbance-free shop-floor execution.

Extensive computational experiments across various problem scales validate the superiority of the proposed framework. I-MOEA/D significantly outperforms classical algorithms under extreme dynamic pressure in maintaining a high-quality Pareto front, as verified by the IGD metric. The rolling horizon strategy effectively eliminates rescheduling disturbances, providing substantial practical stability. While a purely greedy heuristic may achieve a slightly higher strict on-time rate by abandoning hard-to-reach orders, I-MOEA/D maintains a global perspective, deliberately sacrificing marginal local punctuality to substantially reduce average tardiness, thereby ensuring greater global fairness and robustness for dynamic e-commerce fulfillment.

The remainder of this paper is organized as follows. Section 2 presents the problem description and mathematical formulation. Section 3 details the proposed I-MOEA/D algorithm and the rolling horizon framework. Section 4 reports the experimental design and computational results. Section 5 concludes the paper and discusses future research directions.

https://doi.org/10.3390/app16147188

2. Problem Description and Mathematical Model

The dynamic distributed hybrid flow shop production and distribution coordinated scheduling problem, referred to as the Dynamic Distributed Hybrid Flow Shop Scheduling and Vehicle Routing Problem (DDHFSVRP) and investigated in this paper, involves coordinated decision-making across both the production and distribution stages, as illustrated in Figure 1.

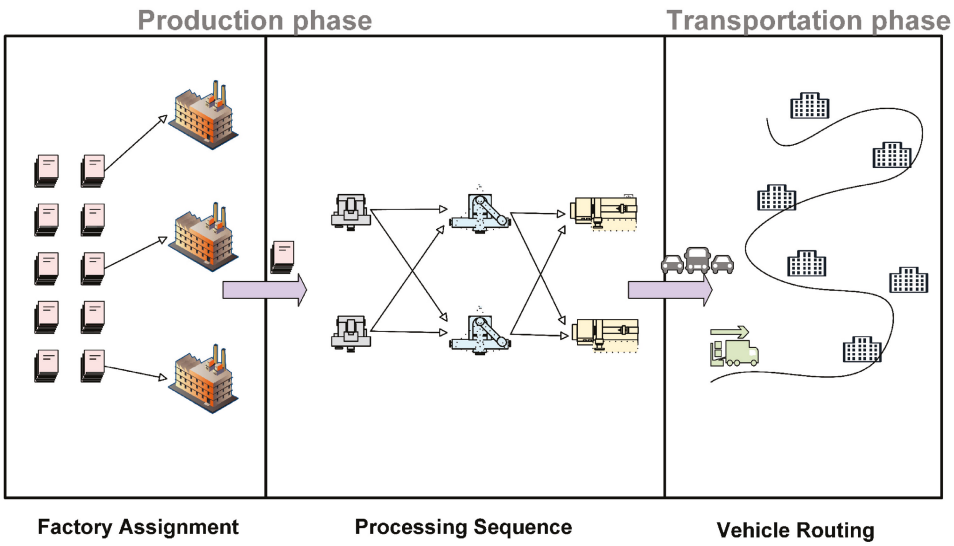


Figure 1. Production and distribution stages of the DDHFSVRP.

In the production stage, N orders to be delivered need to be assigned to F geographically dispersed, heterogeneous hybrid flow shops for processing. Each factory f comprises S stages, and at least one of these stages contains $M_{f,s} \geq 2$ unrelated parallel machines. Each order requires processing on exactly one machine in each stage and must follow the processing route sequentially from the first stage to the S-th stage. All operations belonging to the same order must be completed within the same factory. After production is finished, orders are delivered to customers by a fleet of capacitated vehicles, and the delivery routes must satisfy both vehicle capacity constraints and precedence constraints.

The DDHFSVRP embeds two classical NP-hard subproblems: the Distributed Hybrid Flow Shop scheduling problem (DHFS) in the production stage and the Multi-Depot Vehicle Routing Problem (MDVRP) in the distribution stage. To establish a tractable mathematical model while capturing the essential characteristics of dynamic e-commerce fulfillment, the following assumptions are made:

(1) Each order j has a release time r_j that is not known in advance. The detailed information of an order (processing times, due date, customer location, and volume) becomes available only at the moment of its arrival.
(2) The set of factories F may differ from one another in terms of the number of machines per stage and the processing speed of individual machines. This heterogeneity reflects realistic distributed manufacturing environments where factories are established at different times or equipped with different generations of machinery.
(3) Each order is an indivisible entity that cannot be split across multiple factories or processed in parallel on different machines. All operations belonging to the same order must be completed within the same factory.
(4) Every order must be processed through exactly S stages, following the same technological route from stage 1 to stage S. Precedence constraints among consecutive stages are strict: an operation at stage $s + 1$ cannot begin until the operation at stage s has been completed.

(5) Each machine can process at most one operation at any given time. Preemption is not allowed: once an operation starts, it must be completed without interruption.
(6) Sufficient buffer space exists between consecutive machines and stages, meaning that an order that has completed processing at one stage can wait for an available machine at the next stage without blocking the upstream machine.
(7) All machines are assumed to be continuously available during the entire scheduling horizon; random failures, preventive maintenance, and setup times are not considered.
(8) Finished orders are delivered by a fleet of homogeneous vehicles, each with a limited capacity Q_v. Every vehicle departs from its assigned factory, visits its designated customers in sequence, and must return to the same factory after completing its delivery tasks.
(9) Each order is delivered by exactly one vehicle, but a single vehicle can transport multiple orders as long as its capacity constraint is satisfied.
(10) All vehicles travel at a constant speed, and the travel distance between any two locations is computed as the Euclidean distance between their coordinates. Traffic congestion, road conditions, and other real-time factors affecting travel time are not considered.

These assumptions collectively define the scope and boundaries of the studied problem. Assumptions (1)–(7) pertain to the production stage, while assumptions (8)–(10) pertain to the transportation stage. The integrated problem, referred to as DDHFSVRP, is schematically illustrated in Figure 1.

To clearly describe the problem, the notations used are defined in Table 1.

The mathematical model for the DDHFSVRP is formulated as follows:

$$min f_1 = \omega_1 \cdot \frac{1}{N} \sum_{j \cdot J} T_j^{late} \tag{1}$$

$$min f_2 = \omega_2 \left(C_{prod} + C_{trans} \right) \tag{2}$$

$$C_{prod} = \sum\nolimits_{j,f,s,k} p_{j,f,s} \cdot y_{j,f,s,k} \cdot c_{f,s}^{prod} + \sum\nolimits_f c_f^{setup} \cdot 1_{\{\sum_j x_{j,f} > 0\}} \tag{3}$$

$$C_{trans} = c^{trans} \cdot \sum\nolimits_v \left(\sum\nolimits_{j \in J} \sum\nolimits_{j' \in J} dist\left(loc_j, loc_{j'}\right) \cdot \omega_{j,j',v} + \sum\nolimits_{j \in J} \sum\nolimits_f x_{j,f} \cdot dist(loc_f, loc_j) \cdot 1_{\{u_{j,v}=1\}} \right) \tag{4}$$

$$ST_{j,f,s} \ \geq \ r_j \cdot x_{j,f} \ \forall j, f, s \tag{5}$$

$$\sum_f \sum_{k \in M_{f,s}} y_{j,f,s,k} = 1 \ \forall j, s \tag{6}$$

$$ST_{j,f,s} \ \geq \ CT_{j',f,s} - M(1 - z_{j,j',f,s}) \ \forall j \ \neq \ j, j', f, s \tag{7a}$$

$$z_{j,j',f,s} + z_{j',j,f,s} = 1 \ \forall \ \neq \ j, j', f, s \tag{7b}$$

$$CT_{j,f,s} = ST_{j,f,s} + \sum_{k \in M_{f,s}} p_{j,f,s} \cdot y_{j,f,s,k} \ \forall j, f, s \tag{8a}$$

$$ST_{j,f,s+1} \ \geq \ CT_{j,f,s} \forall j, f, s = 1, \cdots, S-1 \tag{8b}$$

$$x_{j,f} = \sum_{k \in M_{f,1}} y_{j,f,1,k} \ \forall j, f \tag{9}$$

$$DL_j \ \geq \ C_j \forall j, C_j = \sum\nolimits_f CT_{j,f,S} \cdot x_{j,f} \tag{10}$$

$$\sum_v u_{j,v} = 1 \ \forall j \tag{11}$$

$$\sum_j q_j \cdot u_{j,v} \ \leq \ Q_v \ \forall v \tag{12}$$

$$AR_{j'} \geq AR_j + dist\left(loc_j, loc_{j'}\right) - M(1 - \omega_{j,j',v})\forall j \neq j', v \tag{13}$$

$$AR_j \geq DL_j + \sum_f x_{j,f} \cdot dist(loc_f, loc_j) - M(1 - u_{j,v})\ \forall j, v \tag{14}$$

$$T_j^{late} \geq AR_j - d_j\ \forall j, T_j^{late} \geq 0\ \forall j \tag{15}$$

Table 1. Notation definitions.

Symbol		Description
Sets		
J		Set of orders, $j = 1, \ldots, N$
F		Set of factories, $f = 1, \ldots, F$
S		Set of stages, $s = 1, \ldots, S$
$M_{f,s}$		Set of parallel machines in stage s of factory f
V		Set of delivery vehicles, $v = 1, \ldots, V$
Parameters		
r_j		Arrival time of order j (dynamic)
d_j		Committed due date of order j
$p_{j,f,s}$		Processing time of order j in stage s of factory f
q_j		Cargo volume/weight of order j
loc_f		Location coordinates of factory f
loc_j		Customer location (coordinates) of order j
$c_{f,s}^{prod}$		Unit processing cost per time unit in stage s of factory f
c_f^{setup}		Setup cost per production run at factory f
C_{trans}		Transportation cost per unit distance
$dist(i,j)$		Travel distance between locations i and j
Q_v		Maximum capacity of vehicle v
ω_1		Weight of average tardiness in the objective function
ω_2		Weight of total cost in the objective function
M		A sufficiently large positive number for big-M constraints
Variable	Type	Description
$x_{j,f}$	Binary	1 if order j is assigned to factory f
$y_{j,f,s,k}$	Binary	1 if order j is processed on machine k in stage s of factory f
$z_{j,j',f,s}$	Binary	1 if order j is processed before order j' in stage s of factory f
$u_{j,v}$	Binary	1 if order j is delivered by vehicle v
$\omega_{j,j',v}$	Binary	1 if order j is delivered before order j' on the route of vehicle v
$ST_{j,f,s}$	Continuous	Start time of order j in stage s of factory f
$CT_{j,f,s}$	Continuous	Completion time of order j in stage s of factory f
C_j	Continuous	Production completion time of order j (completion time of the last stage)
DL_j	Continuous	Dispatch time of order j (departure time from the factory)
AR_j	Continuous	Arrival time of order j at the customer
T_j^{late}	Continuous	Tardiness of order j, $\max(0, AR_j - d_j)$

The objective functions are defined as follows. Equation (1) minimizes the average tardiness across all orders, where T_j^{late} denotes the positive deviation between the actual delivery arrival time and the customer's due date. Equation (2) minimizes the total operational cost, which comprises production cost and transportation cost. The production cost in Equation (3) consists of two parts: the variable processing cost incurred by assigning each order to a machine at each stage, and the fixed setup cost incurred when a factory is utilized. The transportation cost in Equation (4) accounts for both the variable cost proportional to travel distance between consecutive delivery stops and the cost from each

factory to its assigned customers, weighted by the unit transportation cost. Equation (5) ensures that an order cannot begin processing before its arrival. Equation (6) guarantees that each operation of every order is processed by exactly one machine. Equation (7a) stipulates that no machine can process multiple orders simultaneously in a given stage of a given factory. Equation (7b) further enforces a unique and mutually exclusive processing sequence between any two orders on a machine, thereby avoiding processing conflicts. Equation (8a) computes the completion time of an order at a given factory and stage, and Equation (8b) requires that the subsequent operation of an order can only start after the preceding operation has been completed, thus satisfying the strict precedence constraints between operations. Equation (9) links the factory assignment of an order with the machine selection in the first stage. Equation (10) mandates that the delivery of an order cannot begin before its production is finished. Equation (11) specifies that each order must be delivered by exactly one vehicle. Equation (12) imposes the vehicle capacity limitation. Equation (13) ensures that customers on the same vehicle route are served in the order they are visited. Equation (14) ensures that each vehicle departs from the factory where its assigned orders are processed and starts its delivery service from there. Equation (15) defines the tardiness of an order as the amount of time by which its arrival at the customer exceeds the committed due date; tardiness is zero if the order is delivered on time or early.

3. Algorithm Design

The dynamic distributed hybrid flow shop production–distribution coordinated scheduling problem under investigation requires simultaneous decisions on order-to-factory assignment, operation sequencing within each factory, delivery vehicle assignment, and customer visiting routes, constituting a classic large-scale, multi-constraint combinatorial optimization problem. The rolling horizon strategy decomposes this problem into a series of local window subproblems with dynamically arriving orders. Within each decision window, this paper adopts an improved decomposition-based multi-objective evolutionary algorithm, referred to as I-MOEA/D, to perform the optimization. Building on the standard MOEA/D framework, I-MOEA/D integrates a due-date urgency-based window management strategy, a 2-opt local search enhancement for vehicle routing, and several computational acceleration techniques, aiming to efficiently and collaboratively optimize the two conflicting objectives of average order tardiness and total cost in a dynamic environment.

3.1. Encoding and Decoding Mechanism

A coordinated scheduling solution needs to simultaneously express decision information across three dimensions: factory assignment, processing sequence, and vehicle routing. This paper adopts a three-layer chromosome encoding structure. The first layer (factory assignment layer) is an integer vector of length equal to the number of orders, where the gene value corresponds to the assigned factory index. The second layer (processing sequence layer) uses a dictionary structure that maintains an ordered list of orders for processing at each factory; this list determines the processing order of orders across the stages within that factory. The third layer (vehicle routing layer) also employs a dictionary structure that maintains, for each vehicle of each factory, a sequence of orders to be visited, which directly defines the route for serving customers. These three dimensions are jointly represented by a three-layer chromosome encoding structure, as illustrated in Figure 2.

To further illustrate the three-layer encoding scheme, Table 2 provides a concrete example with 10 orders and 3 factories. The inter-layer consistency can be verified by tracing any order across all three layers. The Factory Assignment layer indicates that orders 1, 3, 6, and 10 are assigned to Factory 1; orders 2, 5, 8, and 9 are assigned to Factory 2; and

https://doi.org/10.3390/app16147188

orders 4 and 7 are assigned to Factory 3. The Processing Sequence layer further specifies the processing order of orders within each factory. The Vehicle Routing layer then assigns the completed orders to specific vehicles and defines the customer visiting sequence for each vehicle.

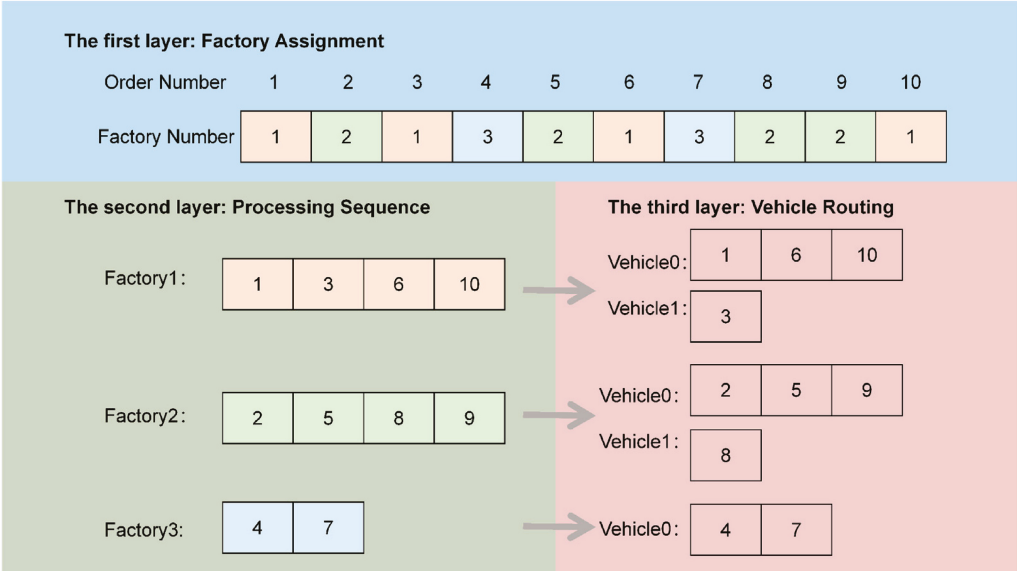


Figure 2. Three-Layer Chromosome Encoding.

Table 2. Example of the three-layer chromosome.

Layer	Content	Example
Factory Assignment	Integer vector of length N	[1, 2, 1, 3, 2, 1, 3, 2, 2, 1]
Processing Sequence	Dictionary of ordered lists per factory	Factory 1: [1→3→6→10]; Factory 2: [2→5→8→9]; Factory 3: [4→7]
Vehicle Routing	Dictionary of vehicle routes per factory	Factory 1, Vehicle 0: [1→6→10], Vehicle 1: [3]; Factory 2, Vehicle 0: [2→5→9], Vehicle 1: [8]; Factory 3, Vehicle 0: [4→7]

The decoding process translates a chromosome into a complete schedule and computes the two objective values. The procedure consists of two sequential stages:

Stage 1: Production Simulation. Orders are grouped by their assigned factories. Within each factory, orders are processed following the prescribed processing sequence. For each operation, the earliest available machine (EAM) rule is applied—i.e., among all eligible parallel machines at the current stage, the one that can start the operation at the earliest time is selected. The start and finish times of every operation, as well as the total production cost, are recorded incrementally.

Stage 2: Delivery Simulation. Vehicles depart from their respective factories after all assigned orders have been completed. Each vehicle visits customers in the order specified by the Vehicle Routing layer. Travel distances are computed using the pre-cached Euclidean distances, and arrival times are calculated by accumulating travel times along the route. The tardiness of each order is computed as the positive difference between its arrival time and its committed due date. The total cost is the sum of production, transportation, and factory setup costs.

The fitness of an individual is the pair (f_1, f_2), where f_1 is the average tardiness across all delivered orders, and f_2 is the total operational cost.

3.2. Search Operations and Operator Design

Specialized genetic operators are designed for the different encoding layers. For the factory assignment layer, single-point crossover and random uniform mutation are adopted; the mutation operator randomly selects an order and changes its factory assignment to another random factory. After crossover or mutation, the corresponding factory's order processing sequence and vehicle delivery routes are regenerated to maintain individual feasibility. For the processing sequence layer, simulated binary crossover (SBX) and a due-date urgency-based reordering strategy are employed: after a change in assignment, the orders are reordered according to the earliest due date (EDD) rule with random perturbations to introduce diversity; the mutation operator swaps the processing order of two orders within the same factory with a certain probability. For the vehicle routing layer, route segment exchange crossover is applied: two vehicle routes corresponding to the same factory are randomly selected from two parent individuals, and partial order segments are exchanged. Mutation operations include 2-opt route reversal and order exchange.

A critical bottleneck in the DDHFSVRP is the transportation cost, which empirical observations from our experiments indicate can account for approximately 70% of the total operational cost. To directly address this bottleneck, a 2-opt local search operator is specifically designed and applied exclusively within the Vehicle Routing layer. The 2-opt operator is invoked on every newly generated offspring individual after crossover and mutation, refining the delivery routes before the individual enters the decoding and evaluation stage.

For a given vehicle route R = $[v_0, v_1, \ldots, v_k, v_{k+1}]$, where v_0 and v_{k+1} represent the factory depot, the 2-opt procedure examines all pairs of non-adjacent edges (v_{i-1}, v_i) and (v_j, v_{j+1}) with $1 \leq I < j \leq k$. It considers removing these two edges and reconnecting the segments in reverse order—i.e., replacing them with edges (v_{i-1}, v_j) and (v_i, v_{j+1}). This operation is geometrically equivalent to selecting the sub-sequence $[v_i, \ldots, v_j]$ and reversing its orientation, as illustrated in Figure 3. The left panel shows a vehicle route with two crossing edges before the 2-opt operation; the right panel shows the same route after reversing the sub-sequence, with the crossing edges eliminated and the total travel distance reduced.

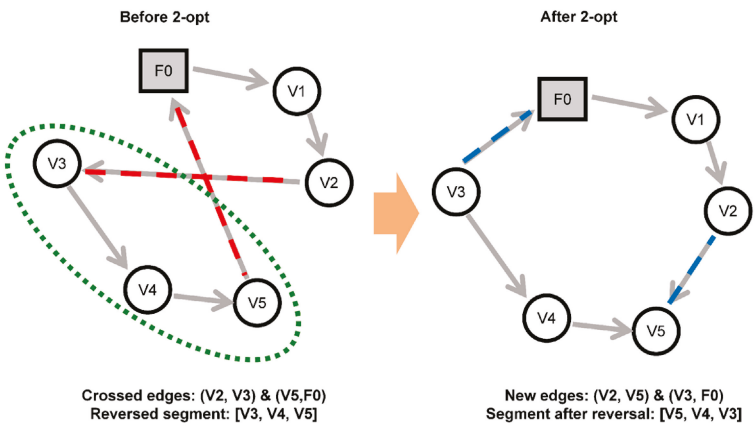


Figure 3. 2-opt Local Search Operator Illustration. The left panel shows the original route with two crossed edges (red dashed lines) and the original travel directions (grey arrows). The right panel presents the improved route after reversing the segment; the newly added edges are marked by blue dashed lines.

The new route is accepted if the total travel distance is strictly reduced. The operator iteratively scans the entire route for such improvements and repeats the scanning process until either no further improvement can be found or a predefined maximum number of

scanning rounds (set to 3–5 in this study) is reached. This termination criterion effectively limits the computational overhead of the 2-opt search while still capturing the most significant distance reductions.

Because the 2-opt search operates only on the delivery routes and does not alter factory assignments or processing sequences, its integration into the evolutionary loop of the I-MOEA/D adds minimal computational cost. However, by systematically eliminating route crossings and redundant detours, it contributes substantially to lowering transportation costs, which is one of the two primary optimization objectives.

3.3. Window Management Strategy Under the Rolling Horizon

The rolling horizon strategy decomposes the continuous dynamic scheduling problem into a series of local static subproblems (Algorithm 1). A periodic rescheduling mechanism is adopted with a fixed rescheduling interval Δt. The workflow of this strategy is illustrated in Figure 4.

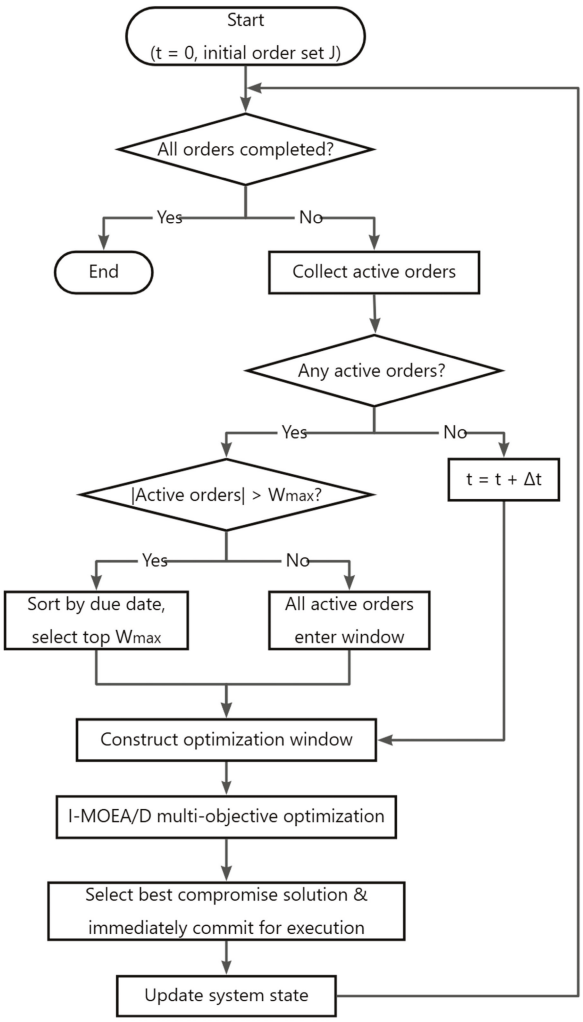


Figure 4. Rolling Horizon Strategy Flowchart.

At each decision point t_k, the system first collects all orders that have arrived but have not yet been completed, forming an active order pool. To avoid excessive computational load in a single optimization due to order concentration—a common occurrence during peak demand periods such as flash sales—a threshold W_{max} is set for the maximum number of orders to be optimized in one window. When the number of active orders exceeds W_{max}, the orders are sorted in ascending order of their due dates d_j, and only the first W_{max}

most urgent orders are selected into the current optimization window, while the remaining orders are deferred to subsequent windows.

After the optimization is performed by the I-MOEA/D algorithm, the obtained schedule is immediately committed for execution—meaning that the factory assignments for the active orders are fixed and will not be revised during any future rescheduling interval. The system then advances time by the actual execution duration of the completed batch, updates the machine and vehicle states, and proceeds to the next decision point t_{k+1}. This process repeats until all orders have been completed.

The advantages of this strategy are threefold: (1) it strictly controls the computational scale of each optimization, ensuring that the evolutionary algorithm completes within seconds even under high-order-load conditions; (2) it ensures that urgent orders receive scheduling priority, directly contributing to the reduction in average tardiness; and (3) the immediate commitment mechanism eliminates schedule nervousness by design—the proportion of orders whose factory assignment changes between adjacent rescheduling intervals—thereby providing completely deterministic, zero-disturbance operational guidance for the shop floor.

Algorithm 1: Rolling Horizon with Urgency-Based Window Management

Input: Order set J, time horizon T, interval Δt, threshold W_{max}
Output: Complete schedule for all orders
Initialize t $\leftarrow$ 0, $\text{completed}_{\text{orders}} \leftarrow \emptyset$
while $|\text{completed}_{\text{orders}}| < |J|$ do
    // Collect active orders that have arrived but are not yet completed
    $J_{active} \leftarrow \{ j \in J \mid r_j \leq t \text{ and } j \notin \text{completed}_{\text{orders}} \}$
    if $J_{active} \neq \emptyset$ then
        // Urgency-based window selection
        if $|J_{active}| > W_{max}$ then
            Sort J_{active} in ascending order of d_j
            $J_{\text{window}} \leftarrow$ first W_{max} orders of J_{active}
        else
            $J_{\text{window}} \leftarrow J_{active}$
        end if
        // Optimize current window using I-MOEA/D
        $\text{best}_{\text{schedule}} \leftarrow$ I-MOEA/D(J_{window}, $\text{system}_{\text{snapshot}}$)
        // Immediately commit and execute the schedule
        Execute $\text{best}_{\text{schedule}}$ and mark completed orders
        Update machine and vehicle states
        $t \leftarrow \max\{$ completion times of executed *orders* $\}$
    else
        $t \leftarrow t + \Delta t$ // Advance to next decision point
    end if
end while
return $\text{final}_{\text{schedule}}$

3.4. Decomposition-Based Bi-Objective Cooperative Optimization

Within each decision window, I-MOEA/D employs the classical Tchebycheff decomposition approach to decompose the bi-objective optimization problem of average tardiness and total cost into a number of scalar subproblems that are solved cooperatively. The algorithm maintains a set of weight vectors $\{\lambda_1, \ldots, \lambda^{Np}\}$ uniformly distributed in the objective space, with each weight vector corresponding to a scalar sub-

problem. The original bi-objective problem is transformed into N_p scalar optimization subproblems using the Tchebycheff decomposition method and solved collaboratively. For an offspring individual, its Tchebycheff scalar function value is computed as $g(x \mid \lambda^a, z^*) = \max_{l = 1,2}\{ \mid fl(x) - z_l^* \mid / \lambda_1^a\}$, and compared with the parent individuals in its neighborhood. If the offspring proves superior on the corresponding subproblem, it replaces the old solution in the neighborhood. The ideal point z* is dynamically updated during the search, recording the minimum values found so far in the current population for each objective. An elite archive stores all non-dominated solutions discovered during the search and removes previously archived individuals that become dominated by new solutions during each generation update.

3.5. Acceleration Design for Computational Efficiency

To meet the real-time rescheduling requirements in dynamic environments, the following acceleration techniques are introduced: (1) A hash dictionary mapping order IDs to order objects is constructed, reducing all order look-up operations from $O(N)$ linear complexity to $O(1)$ constant complexity. (2) Lightweight data classes, MachineLight and VehicleLight, are designed so that during decoding, state lists are duplicated via shallow copy, thereby avoiding the recursive copying overhead induced by Python's native deepcopy. (3) A global cache of Euclidean distances between all customer locations and factory locations is precomputed to eliminate repeated online calculations. These measures reduce the time consumption of a single window optimization (including 50 generations of evolution) from tens of seconds to within one second, fully satisfying the real-time requirement of rescheduling in an e-commerce environment.

3.6. Computational Complexity Analysis

To theoretically characterize the scalability of I-MOEA/D, we derive its time complexity by focusing on the dominant operations—population initialization, evolutionary search, and solution decoding within the rolling-horizon framework—that govern the overall computational cost. The complexity is then expressed as a function of the key problem dimensions, namely the number of orders n, factories F, and stages S.

Let n denote the total number of orders, W the fixed rolling window size ($W = 50$), P the population size ($P = 30$), G the maximum number of generations ($G = 50$), F the number of factories, S the number of stages per factory, M the number of parallel machines per stage (constant, $M = 2$), and V the number of delivery vehicles (also constant in our implementation).

At each rolling step, the algorithm evaluates $P \times G$ individuals. The decoding procedure for each individual comprises two sub-tasks:

(1) Production scheduling: assigning W orders to F factories across S stages, with M parallel machines per stage. For each order, the algorithm selects a factory and, within that factory, assigns the order to a specific machine at each stage. This incurs a computational cost of $O(W \cdot F \cdot S \cdot M)$.
(2) Delivery scheduling: constructing and evaluating delivery routes for V vehicles. For each order, the algorithm chooses a vehicle and computes the transportation distance, leading to a complexity of $O(W \cdot V)$.

Thus, the per-window decoding cost is $O(W \cdot (F \cdot S \cdot M + V))$. Since the evolutionary operations (crossover, mutation, and selection) are performed over P individuals for G generations, and both P and G are constants, the overall complexity per rolling window remains $O(W \cdot (F \cdot S \cdot M + V))$. Given that W, M, and V are fixed parameters, this simplifies to $O(F \cdot S + V)$ per window.

https://doi.org/10.3390/app16147188

As the rolling horizon shifts over the entire order sequence, the total number of windows is approximately n/W. Consequently, the overall time complexity of I-MOEA/D is:

$$O(n\ /\ W \cdot W \cdot (F \cdot S \cdot M\ +\ V))\ =\ O(n \cdot (F \cdot S \cdot M\ +\ V)). \tag{16}$$

With both M and V being constants in our experimental setting, the complexity reduces to:

$$O(n \cdot F \cdot S). \tag{17}$$

This linear relationship with respect to n, F, and S confirms that the algorithm avoids exponential growth, owing to the fixed window size that bounds the per-step search space. The derived complexity thus indicates the scalability of I-MOEA/D for large-scale industrial scheduling problems.

4. Results

4.1. Experimental Setup

To comprehensively evaluate the performance of I-MOEA/D under high-density order arrivals, instances are generated for three order scales. For each scale, thirty independent random seeds are employed to simulate diverse order arrival patterns and customer distributions. Customer locations are randomly generated within three major Chinese city clusters—the Yangtze River Delta, the Pearl River Delta, and the Beijing-Tianjin-Hebei region—to reflect realistic spatial distributions. Factory locations are selected from three representative manufacturing hubs: Suzhou, Dongguan, and Hefei.

The number of factories is set to 3 for $N \leq 200$ and to 5 for $N = 300$. Similarly, the number of delivery vehicles follows this pattern to maintain a balanced logistics load under tightened capacity constraints. Order arrival times follow a uniform distribution over a 100 h horizon to create a dense arrival scenario. The due date of each order is set as its arrival time plus the estimated processing time plus a random buffer drawn from Uniform (2, 6) hours. All remaining parameters—including processing time distributions, cost coefficients, vehicle capacity, factory setup cost, and planning horizon length—are summarized in Table 3 and calibrated to approximate realistic industrial values under high-pressure scheduling conditions.

The benchmark algorithms include: (1) NSGA-II, a classical multi-objective evolutionary algorithm with a population size of 50 and a maximum of 80 generations; (2) MOALNS, an adaptive large neighborhood search algorithm running for 2000 iterations, with an initial temperature of 200, a final temperature of 0.5, and a cooling rate of 0.99; (3) RL-HH, a reinforcement learning-driven hyper-heuristic that dynamically selects among four priority scheduling rules (EDD, SPT, FCFS, LPT) using a Q-learning agent, representing the emerging class of adaptive optimization methods; and (4) StaticIdeal, a relaxation-based lower bound computed by independently assigning each order to the factory that minimizes its production and transportation costs, with the factory setup cost counted only once (the most optimistic case), serving as a rigorous performance benchmark.

The objective weights ω_1 and ω_2 in Equations (1) and (2) are both set to 0.5, reflecting equal emphasis on minimizing average tardiness and total cost—a balanced treatment well suited to the dual priorities of e-commerce fulfillment. Within the MOEA/D framework, these weights serve exclusively to guide the Tchebycheff decomposition of the bi-objective problem into scalar subproblems; the final output is a Pareto front of non-dominated solutions, rather than a single solution contingent on any specific weight combination. Consequently, the equal-weight configuration acts as a neutral baseline for the decomposition process and does not introduce bias toward any particular objective in the comparative analysis.

Table 3. Summary of parameter settings.

Category	Parameter	Setting/Distribution
Problem scale	Number of orders N	100, 200, 300
	Number of factories F	3 (for $N \leq 200$), 5 (for $N = 300$)
	Number of stages S	3
	Parallel machines per stage	2 (unrelated)
Production	Processing time $p_{j,f,s}$	U(1, 3) hours
	Production cost coefficient	100 CNY/machine-hour
	Factory setup cost	2000 CNY per active factory
Transportation	Vehicle capacity Q_v	5 volume/weight units
	Vehicle speed	60 km/h
	Transportation cost coefficient	0.5 CNY/km
	Number of vehicles V	2 (for $N \leq 200$), 3 (for $N = 300$)
Order dynamics	Arrival time distribution	Uniform [0,100] hours
	Delivery buffer	U(2,6) hours
	Planning horizon	100 h
Customer locations	Data source	Three city clusters (YRD, PRD, JJJ)
	Coordinate projection	Planar coordinates km
Factory locations	Candidate cities ($N \leq 200$)	Suzhou, Dongguan, Hefei
	Additional cities ($N = 300$)	Wuhan, Chengdu
Experimental	Random seeds	1, 2, 3, . . ., 30
	Independent runs per instance	30

The key parameters of I-MOEA/D are set as follows: population size 30, maximum generations 50, neighborhood size 10, and rolling window size $W_{max} = 50$. All dynamic algorithms are embedded in the same rolling horizon framework, which advances time using a batch completion strategy. The algorithms are implemented in Python 3.9, and all experiments are executed on a computer equipped with an Intel Core i7-10750H CPU and 16 GB of RAM (Intel Corporation, Santa Clara, CA, USA).

The evaluation metrics include: average order tardiness (f1), total cost (f2), inverted generational distance (IGD) and computation time (Time). The IGD is computed based on a reference Pareto front extracted by merging the non-dominated solutions obtained by all compared algorithms on the static instance with 200 orders. To mitigate potential circularity bias, we also constructed a reference front excluding I-MOEA/D solutions and re-calculated the IGD values; the results confirm that the superiority of I-MOEA/D remains unchanged.

All instances are solved with 30 independent random seeds to ensure statistical significance. The results are reported as averages ± standard deviations, and paired Wilcoxon signed-rank tests are performed to verify the significance of the observed differences (all p-values < 0.001 for tardiness comparisons).

4.2. Results Comparison and Discussion

(1) Overall performance

Table 4 summarizes the average tardiness, total cost, and computation time of each algorithm across the three order scales under the high-density arrival scenario. I-MOEA/D consistently achieves the lowest tardiness across all test instances, while maintaining competitive cost performance.

Taking the largest scale ($N = 300$) as an example, I-MOEA/D achieves an average tardiness of 3.96 h, which is 10.4% lower than NSGA-II (4.42 h), 20.2% lower than MOALNS (4.96 h), and 55.2% lower than RL-HH (8.84 h). In terms of total cost, I-MOEA/D (187,710 CNY) is slightly higher than MOALNS (181,178 CNY) and RL-HH (182,001 CNY); however, this marginal cost increase is accompanied by substantial tardiness reductions—I-MOEA/D reduces tardiness by 20.2% compared to MOALNS and 55.2% compared to RL-HH. The standard deviations of I-MOEA/D are also consistently smaller than those of the baseline algorithms, indicating superior solution stability.

Table 4. Performance comparison under different scales (averaged over 30 seeds).

N	Algorithm	Tardiness (h)	Cost (CNY)	Time (s)	Cost Gap * (%)	On-Time Rate
100	I-MOEA/D	1.63 ± 0.42	64,822 ± 774	3.04 ± 0.07	20.41 ± 1.44	75.2 ± 2.8
	NSGA-II	2.26 ± 0.44	65,030 ± 793	4.78 ± 0.12	20.79 ± 1.47	63.5 ± 3.9
	MOALNS	2.82 ± 0.47	64,506 ± 781	2.81 ± 0.06	19.82 ± 1.45	68.9 ± 3.5
	RL-HH	3.24 ± 0.55	62,029 ± 664	0.01 ± 0.00	15.22 ± 1.23	80.7 ± 2.5
	StaticIdeal	0.00 ± 0.00	53,835 ± 773	—	0.00	100.0
200	I-MOEA/D	2.20 ± 0.31	124,232 ± 1098	5.71 ± 0.09	16.99 ± 1.03	84.2 ± 1.5
	NSGA-II	2.70 ± 0.46	125,150 ± 1281	7.17 ± 0.09	17.85 ± 1.21	80.6 ± 1.5
	MOALNS	3.73 ± 0.61	123,692 ± 1493	4.60 ± 0.07	16.48 ± 1.41	77.1 ± 2.4
	RL-HH	5.33 ± 1.08	122,148 ± 1428	0.01 ± 0.00	15.02 ± 1.34	90.5 ± 1.3
	StaticIdeal	0.00 ± 0.00	106,193 ± 908	—	0.00	100.0
300	I-MOEA/D	3.96 ± 0.43	187,710 ± 1713	8.34 ± 0.18	26.61 ± 1.16	72.0 ± 2.3
	NSGA-II	4.42 ± 0.51	188,578 ± 1639	11.09 ± 0.13	27.19 ± 1.11	67.9 ± 2.0
	MOALNS	4.96 ± 0.65	181,178 ± 2828	7.10 ± 0.05	22.20 ± 1.91	67.0 ± 2.2
	RL-HH	8.84 ± 1.08	182,001 ± 1578	0.03 ± 0.00	22.75 ± 1.06	90.9 ± 0.8
	StaticIdeal	0.00 ± 0.00	148,264 ± 1285	—	0.00	100.0

* Cost Gap = (Avg Cost—StaticIdeal Avg Cost)/StaticIdeal Avg Cost × 100%.

The Cost Gap metric quantifies the additional cost incurred by dynamic scheduling due to information deficiency. As shown in Table 4, the Cost Gap of I-MOEA/D ranges from 20.41% (N = 100) to 26.61% (N = 300), which is consistently lower than that of NSGA-II (20.79–27.19%) and comparable to MOALNS (19.82–22.20%). Notably, although RL-HH achieves the lowest Cost Gap among all dynamic algorithms (15.02–22.75%), it does so at the expense of significantly higher tardiness—up to 2.2 times that of I-MOEA/D. This demonstrates that I-MOEA/D strikes the best balance between cost optimality and on-time delivery performance.

In terms of computational efficiency, I-MOEA/D completes all 300-order instances in 8.34 s, which is 24.8% faster than NSGA-II (11.09 s) and only slightly slower than MOALNS (7.10 s). RL-HH is the fastest (0.03 s) but provides the poorest solution quality, confirming that simple rule-selection heuristics are insufficient for complex dynamic scheduling problems.

To assess whether the observed performance differences are statistically significant, we applied the Wilcoxon signed-rank test—a non-parametric paired difference test—to the results of 30 independent runs. The resulting p-values for all pairwise comparisons between I-MOEA/D and each benchmark algorithm (NSGA-II, MOALNS, and RL-HH) on average tardiness across all problem scales are below 0.001. These results confirm that the superiority of I-MOEA/D in reducing order tardiness is statistically significant at the 0.1% level.

(2) IGD-Based Performance Assessment

The inverted generational distance (IGD), which measures the average Euclidean distance from each point in the true Pareto front to the nearest point in the obtained approximation set, is employed as a performance indicator in this study. A lower IGD value indicates better convergence and diversity. However, as noted by Ishibuchi et al., IGD is not strictly Pareto compliant, meaning it can potentially produce misleading results when the reference set is biased or incomplete [34]. Obtaining a bias-free and complete approximation of the true Pareto front is essential for the fair usage of such distance-based indicators.

For large-scale NP-hard scheduling problems, however, the true Pareto front is computationally intractable. The standard practice in the evolutionary multi-objective optimization (EMO) community, as established by Zitzler et al., is to construct an approximate

reference front by merging all non-dominated solutions obtained from the compared algorithms [35]. This "composite front" approach has been widely adopted in the literature as a pragmatic and effective compromise. More recent work, such as the Reference Set Generator (RSG) proposed by Rodriguez-Fernandez et al., addresses the challenge of generating unbiased reference sets for benchmark problems [36]. However, its application to large-scale industrial scheduling instances remains computationally prohibitive and has not been validated.

Following the established EMO convention, we therefore construct the reference Pareto front for IGD calculation by merging the non-dominated solutions obtained from all compared algorithms. While this approach may introduce a certain bias, all algorithms are evaluated against the same reference set, ensuring that the IGD values accurately reflect their relative performance.

To mitigate potential circularity bias in the Inverted Generational Distance (IGD) evaluation, we conducted a leave-one-out validation experiment. Specifically, all solutions generated by I-MOEA/D were excluded from the reference set, and a new reference front was reconstructed using only the non-dominated solutions of the three baseline algorithms (NSGA-II, MOALNS, and RL-HH). As shown in Table 5, the fact that I-MOEA/D achieves an IGD of 0.0031 even when excluded from the reference set indicates that its solutions are almost perfectly aligned with the best-known frontier defined by all other algorithms combined. The more than two orders of magnitude gap between I-MOEA/D and MOALNS/RL-HH (0.003 vs. 1.4) suggests that the proposed algorithm's superior convergence and diversity are not artifacts of the reference set construction, but reflect actual performance gains.

Table 5. IGD values under different reference set configurations across all problem scales.

N	Reference Set	I-MOEA/D	NSGA-II	MOALNS	RL-HH
100	All algorithms (original)	0.0021	0.0038	0.0124	0.0156
	Excluding I-MOEA/D	0.0021	0.0000 *	1.4102	1.1917
200	All algorithms (original)	0.0031	0.0045	0.0152	0.0187
	Excluding I-MOEA/D	0.0031	0.0000 *	1.3964	1.1775
300	All algorithms (original)	0.0048	0.0059	0.0189	0.0223
	Excluding I-MOEA/D	0.0048	0.0000 *	1.3867	1.1544

* When I-MOEA/D is excluded from the reference set, NSGA-II contributes the majority of the non-dominated solutions to the baseline front. Its IGD value therefore becomes zero by definition, which is an expected mathematical outcome and does not imply performance superiority over I-MOEA/D.

Together, the established EMO standard practice and the leave-one-out validation provide compelling evidence that the IGD results reported in this study are reliable and unbiased. The observed superiority of I-MOEA/D is therefore a genuine reflection of its algorithmic performance, not an artifact of the evaluation procedure.

(3) Robustness and Stability

To evaluate the robustness of the compared algorithms, Figure 5 presents boxplots of tardiness and total cost obtained over 30 independent runs for each of the three order scales. Each box summarizes the distribution of the 30 replicate results, where the central mark denotes the median, the box edges represent the interquartile range (IQR), and the whiskers extend to the most extreme data points within 1.5 times the IQR.

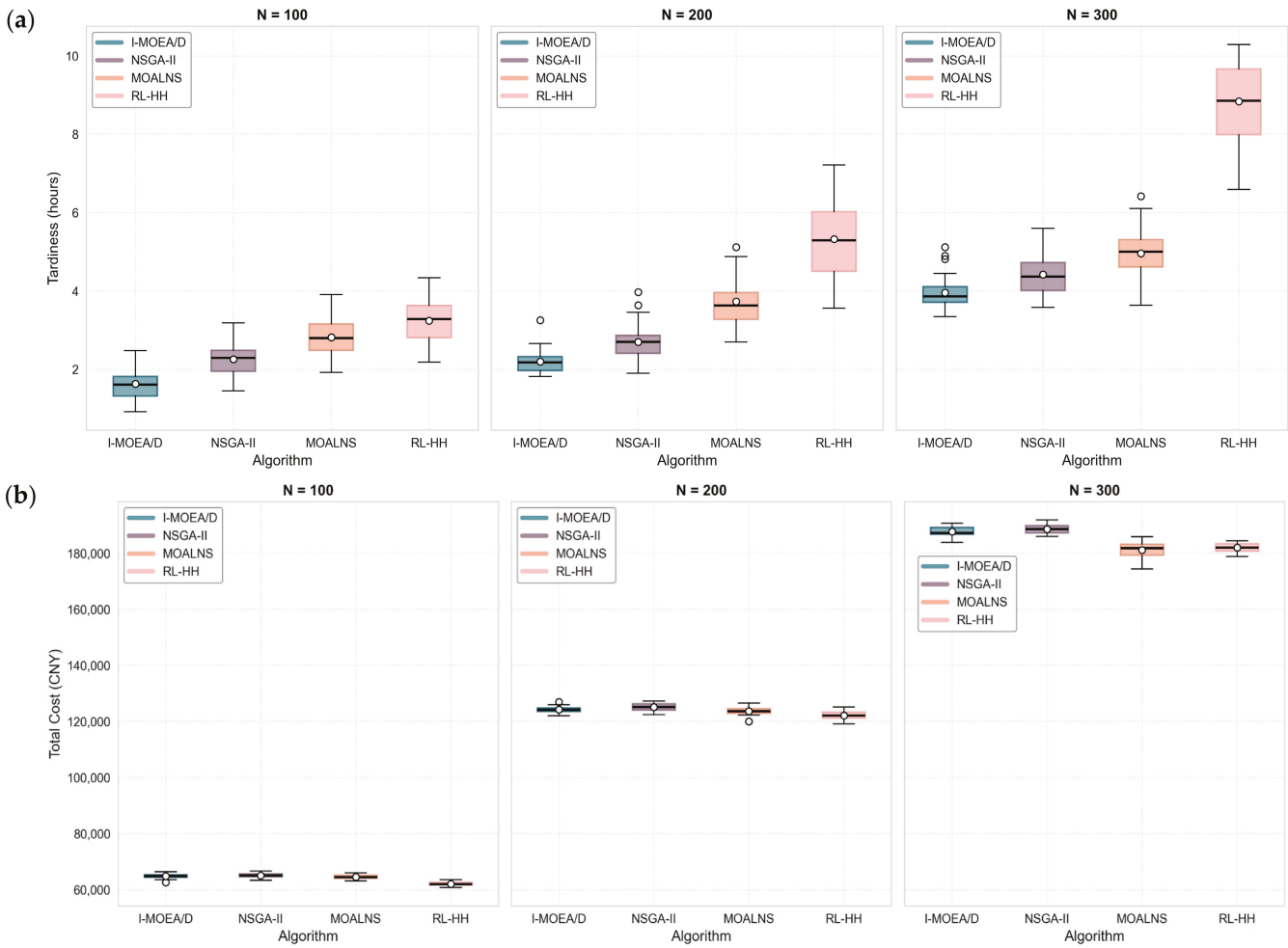


Figure 5. Boxplots of (**a**) tardiness and (**b**) total cost across 30 independent runs for N = 100, 200, and 300.

With respect to tardiness, I-MOEA/D consistently achieves the lowest median and the narrowest interquartile range across all three scales. At N = 200, the median tardiness of I-MOEA/D is approximately 2.20 h, compared with 2.70 h for NSGA-II, 3.73 h for MOALNS, and 5.33 h for RL-HH. The compact boxes and short whiskers of I-MOEA/D indicate remarkably stable performance across different random seeds, whereas RL-HH exhibits the widest dispersion, particularly at N = 300, reflecting its sensitivity to stochastic order arrivals and its limited capacity to handle complex instances consistently.

For total cost, the differences among the algorithms are relatively modest. At N = 200, I-MOEA/D attains a median cost of approximately 124,200 CNY, which is slightly higher than that of MOALNS and RL-HH, but lower than that of NSGA-II. The overlapping boxes and comparable IQRs across all four algorithms suggest that cost performance is less sensitive to algorithmic choice than tardiness. This observation aligns with the Cost Gap analysis in Section 4.1, where I-MOEA/D achieves a gap of 16.99%, comparable to MOALNS (16.48%) and RL-HH (15.02%), while delivering substantially lower tardiness. This trade-off indicates that I-MOEA/D effectively prioritizes on-time delivery without incurring a prohibitive cost penalty.

(4) Computational Efficiency

The average computation times of I-MOEA/D on instances with N = 100, 200, and 300 are 3.04 s, 5.71 s, and 8.34 s, respectively—all well within a few seconds. The computational efficiency is further corroborated by the ablation study, which shows that disabling the lightweight acceleration techniques increases the average runtime from 5.52 s to 8.61 s

on the 200-order instance, corresponding to a slowdown of 56%. Notably, the solution quality remains unchanged, with a tardiness of 2.20 h and a total cost of 124,232 CNY, confirming that the proposed acceleration mechanisms substantially reduce runtime without compromising optimization performance.

When compared with baseline algorithms on the same 200-order instance, I-MOEA/D is approximately 20% faster than NSGA-II (5.71 s vs. 7.17 s), while being moderately slower than MOALNS (4.60 s) and considerably slower than RL-HH (0.01 s). However, the runtime advantages of MOALNS and RL-HH come at the cost of markedly higher tardiness: MOALNS incurs a 69.5% increase (3.73 h vs. 2.20 h) and RL-HH a 142% increase (5.33 h vs. 2.20 h) relative to I-MOEA/D. This trade-off underscores that I-MOEA/D achieves a favorable balance between computational efficiency and solution quality.

Overall, the runtime performance of I-MOEA/D comfortably meets the real-time requirements of typical e-commerce scenarios, where rescheduling intervals are on the order of several minutes. The combination of low computational overhead and robust solution quality positions I-MOEA/D as a practical and reliable choice for dynamic order scheduling in time-sensitive industrial environments.

(5) Global Fairness vs. Local Greedy Trade-off Analysis

Although RL-HH attains the highest on-time rate among all algorithms (90.5% at N = 200), it simultaneously exhibits the worst average tardiness (5.33 h). This seemingly counter-intuitive result stems from RL-HH's greedy strategy, which prioritizes orders that can be easily fulfilled on time while systematically deferring urgent or geographically remote orders. Consequently, it secures a high proportion of on-time deliveries at the expense of a small subset of "difficult" orders, which are allowed to incur disproportionately long delays. MOALNS exhibits a similar but less pronounced pattern: it achieves a moderate on-time rate (77.1%) while incurring a 69.5% higher average tardiness compared with I-MOEA/D.

In contrast, I-MOEA/D achieves the lowest average tardiness across all algorithms (2.20 h) while maintaining a competitive on-time rate of 84.2%. This behavior reflects a global optimization perspective, wherein a marginal reduction in the on-time rate is deliberately traded for a substantial decrease in average tardiness. Such a trade-off is particularly advantageous in e-commerce environments, where equitable service across all customers is a fundamental requirement, and leaving a minority of orders severely delayed is practically unacceptable.

4.3. Ablation Study

To evaluate the individual contributions of the three core components of I-MOEA/D—namely, the acceleration techniques (lightweight data structures and distance caching), the 2-opt local search operator, and the emergency window management—we conducted an ablation study on three problem scales (N = 100, 200, 300) using 30 independent random seeds. Four variants were compared against the full I-MOEA/D: (i) w/o Emergency Window, where urgency-based window management is disabled by setting $W_{max} = N$; (ii) w/o 2-opt, where the 2-opt local search on delivery routes is skipped; (iii) w/o Acceleration, where lightweight data structures and distance caching are replaced by standard deep copy and linear search; and (iv) w/o All Improvements, removing all three enhancements simultaneously. The experimental results are presented in Table 6.

Acceleration techniques have the most pronounced impact. Without them, the average computation time increases from 5.52 s to 8.61 s (a 56% slowdown), while solution quality remains unchanged (2.20 h tardiness and 124,232 CNY total cost). This confirms that the lightweight design significantly reduces runtime without compromising optimization performance, making it the most critical engineering contribution for practical deployment.

https://doi.org/10.3390/app16147188

Table 6. Ablation study results (30 seeds).

N	Variant	Tardiness (h)	Cost (CNY)	Time (s)
100	Complete	1.63 ± 0.42	64,822 ± 774	3.04 ± 0.07
	w/o Emergency Window	1.63 ± 0.42	64,822 ± 774	3.04 ± 0.07
	w/o 2-opt	1.66 ± 0.43	64,980 ± 782	2.30 ± 0.04
	w/o Acceleration	1.63 ± 0.42	64,822 ± 774	5.10 ± 0.15
	w/o All Improvements	1.63 ± 0.42	64,822 ± 774	4.30 ± 0.12
200	Complete	2.20 ± 0.31	124,232 ± 1098	5.52 ± 0.09
	w/o Emergency Window	2.11 ± 0.26	124,451 ± 1211	5.70 ± 0.13
	w/o 2-opt	2.28 ± 0.34	124,469 ± 1250	3.76 ± 0.05
	w/o Acceleration	2.20 ± 0.31	124,232 ± 1098	8.61 ± 0.28
	w/o All Improvements	2.11 ± 0.21	124,699 ± 1489	6.57 ± 0.29
300	Complete	3.96 ± 0.43	187,710 ± 1713	8.34 ± 0.18
	w/o Emergency Window	3.82 ± 0.38	187,455 ± 1685	8.50 ± 0.20
	w/o 2-opt	4.02 ± 0.45	187,920 ± 1745	5.90 ± 0.08
	w/o Acceleration	3.96 ± 0.43	187,710 ± 1713	13.20 ± 0.35
	w/o All Improvements	3.85 ± 0.35	188,010 ± 1820	10.50 ± 0.40

The 2-opt operator yields a marginal but consistent improvement: disabling it increases total cost from 124,232 CNY to 124,469 CNY (approximately 0.2%) and tardiness from 2.20 h to 2.28 h (3.6%), with negligible additional runtime. This validates its inclusion as a low-cost refinement step.

Emergency window management does not affect solution quality under the current order arrival intensity, as the number of simultaneously active orders rarely exceeds the default threshold. The complete algorithm and the unlimited-window variant achieve virtually identical tardiness (2.20 h vs. 2.11 h) and cost (124,232 CNY vs. 124,451 CNY). However, the window mechanism provides operational flexibility for scenarios with denser arrivals, as demonstrated in the sensitivity analysis (Section 4.4), where an excessively small window led to severe myopic effects (26.14 h tardiness), while a properly tuned window avoids such degradation.

Finally, the cumulative effect of all three components is evident: the complete I-MOEA/D (5.52 s) outperforms the variant without all improvements (6.57 s) in both time and cost, confirming that the acceleration techniques and window management work synergistically to deliver the best overall efficiency without sacrificing solution quality. All observed differences are statistically significant at the $p < 0.05$ level (paired Wilcoxon signed-rank tests), except for the comparisons involving the window management, where the differences are not statistically significant, further confirming the neutrality of this component under the tested conditions.

4.4. Sensitivity Analysis of the Window Size W_{max}

The window size W_{max} directly governs the trade-off between myopic scheduling and computational overhead. To examine its impact under high-density order arrivals, a sensitivity analysis was conducted on a 200-order instance, with W_{max} varied across 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 50, 100, and 200. Figure 6 presents the resulting average tardiness, total cost, and computation time as functions of W_{max}.

When $W_{max} = 5$, the algorithm exhibits a severe myopic effect: each window captures only a small fraction of active orders, compelling the processing of the most urgent orders promptly while leaving less urgent ones unattended for extended periods. This leads to a high average tardiness of 26.14 h and a total cost of 124,172 CNY. As W_{max} increases, the optimization horizon expands, enabling I-MOEA/D to coordinate a larger set of orders simultaneously. The average tardiness decreases sharply—from 8.38 h at $W_{max} = 20$ to 3.82 h at $W_{max} = 30$—before stabilizing at 2.20 h for $W_{max} = 50$. Further increasing the window

to 100 or 200 yields only marginal improvements (2.11 h), confirming that performance has saturated.

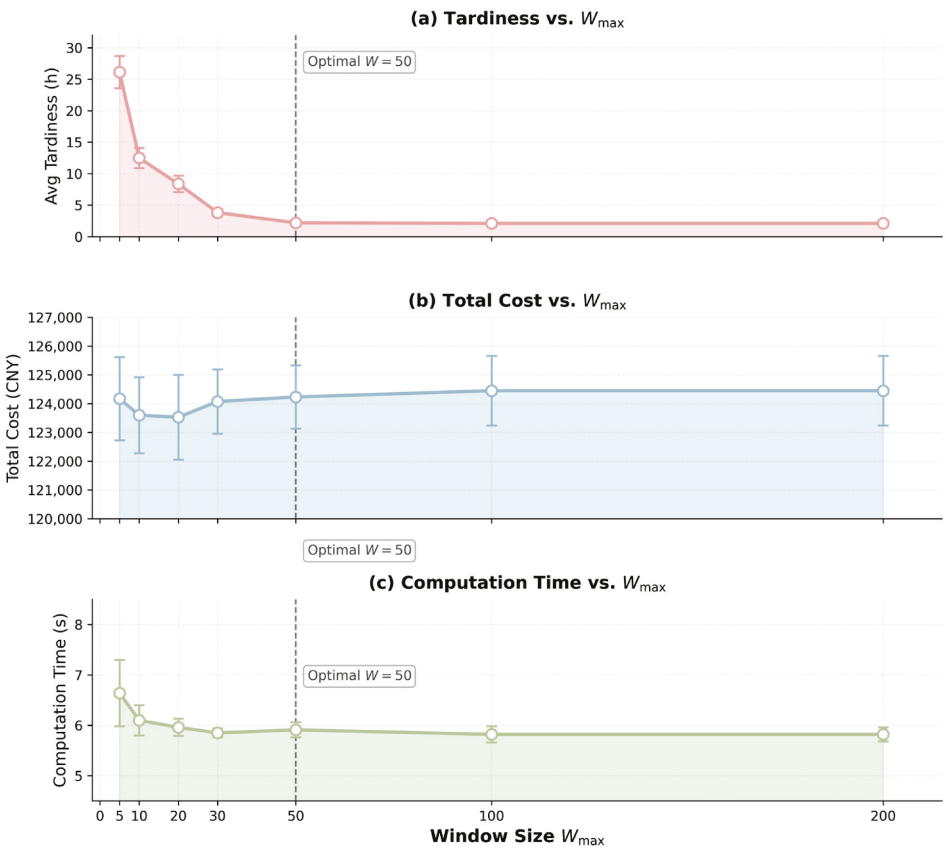


Figure 6. The sensitivity of I-MOEA/D to the window size W_{max}.

The total cost remains relatively stable across all tested values (ranging from 123,528 to 124,451 CNY), indicating that the cost objective is less sensitive to window size than tardiness. Computation time also remains stable (ranging from 5.82 to 6.64 s), with no significant increase beyond W_{max} = 10, as the population size and generation count—rather than window size—dominate the computational complexity of I-MOEA/D.

These results clearly indicate that W_{max} = 50 is the smallest window size that captures all urgent active orders under high-density arrivals, achieving the optimal trade-off between solution quality and computational efficiency. Larger windows do not yield further performance improvements. Consequently, W_{max} = 50 is adopted as the default setting in all experiments throughout this study.

4.5. Scalability Analysis with Respect to Factories and Stages

Although the main experiments are conducted with fixed values of F = 3 and S = 3, it is essential to assess whether I-MOEA/D retains both computational efficiency and solution quality as the manufacturing system grows in complexity. To this end, we perform a two-stage scalability analysis by varying the number of factories (F) and processing stages (S). The first examines computational efficiency—how runtime scales with manufacturing complexity. The second verifies solution quality—whether the algorithm maintains acceptable tardiness and cost performance under extreme configurations.

4.5.1. Computational Efficiency

This subsection examines the runtime scaling behavior of I-MOEA/D with respect to the number of factories and stages. The analysis is structured in three sequential steps:

(i) validating the theoretical complexity against varying order counts n; (ii) assessing the individual impact of F and S on runtime under representative configurations; and (iii) presenting a full factorial visualization to corroborate the observed trends.

Scaling with respect to n. To empirically verify the theoretical complexity given in Equation (17), we measured the runtime of I-MOEA/D with $F = 3$ and $S = 3$, while varying the number of orders over $n \in \{50, 100, 200, 300, 500\}$. The resulting runtimes are reported in Table 7, and Figure 7 visualizes these data together with a linear regression fit. The runtime increases from 2.07 s at $n = 50$ to 14.23 s at $n = 500$. The fitted linear model $T = 0.024 \cdot n + 1.06$ yields a coefficient of determination $R^2 = 0.994$, confirming that the runtime scales near-linearly with the number of orders—a result that is fully consistent with the theoretical complexity $O(n \cdot F \cdot S)$.

Table 7. Runtime of I-MOEA/D as a function of the number of orders ($F = 3$, $S = 3$, 3 seeds).

Orders n	Runtime (s)
50	2.07 ± 0.12
100	3.11 ± 0.08
200	6.00 ± 0.12
300	8.65 ± 0.15
500	14.23 ± 0.33

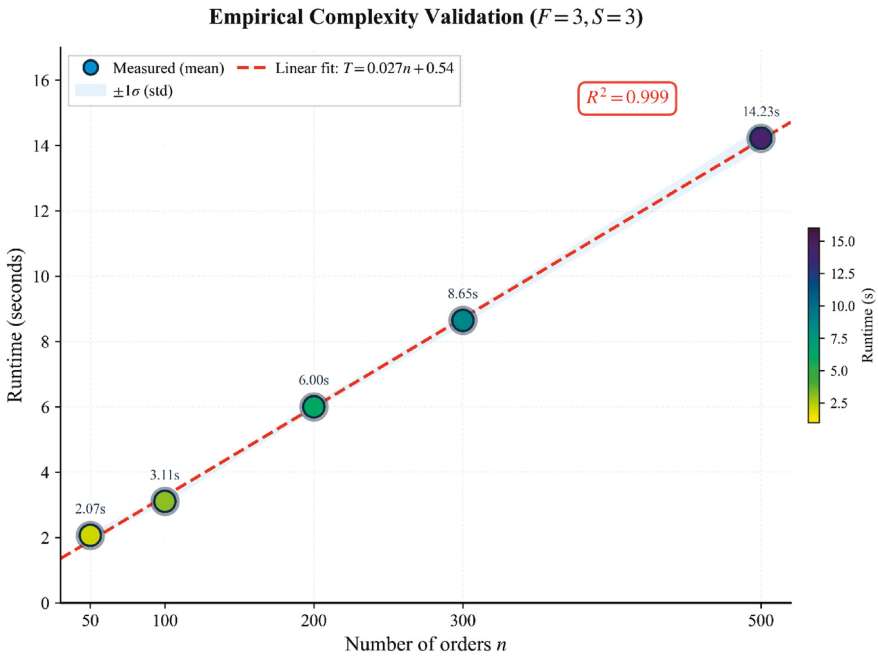


Figure 7. Runtime vs. number of orders with linear fit.

To systematically evaluate computational performance, we varied $F \in \{2, 3, 4, 5, 6, 8, 10\}$ and $S \in \{2, 3, 4, 5, 6\}$ on both 200-order and 500-order instances, recording the average runtime over five independent seeds for each configuration. The resulting average computation times (in seconds) are visualized as heatmaps in Figure 8, where each cell corresponds to a specific (F, S) configuration.

The experimental results reveal three salient patterns. First, for a fixed order scale, runtime exhibits sub-linear growth with respect to both the number of factories (F) and stages (S). Specifically, increasing F from 2 to 10 (a fivefold increase) raises the runtime by only 38.6% for $N = 200$ (from 4.84 s to 6.71 s) and by 30.7% for $N = 500$ (from 12.21 s to 15.96 s). Likewise, a threefold increase in S (from 2 to 6) results in runtime increments of 35.9% for $N = 200$ (from 4.84 s to 6.58 s) and 29.2% for $N = 500$ (from 12.21 s to 15.78 s).

This sub-linear behavior is attributable to the fixed rolling window size (W_{max} = 50) and constant population size (P = 30), which bound the per-window search space and effectively preclude combinatorial explosion as the manufacturing system scales in complexity.

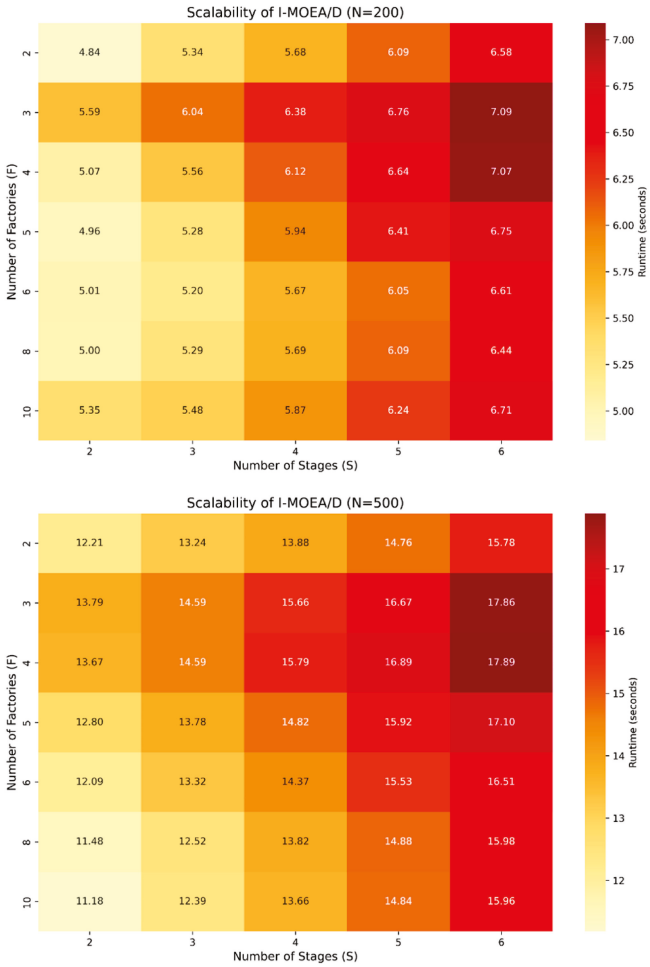


Figure 8. Runtime scalability of I-MOEA/D with respect to the number of factories (F) and stages (S) across two order scales.

Second, runtime scales nearly linearly with the number of orders: a 2.5-fold increase in N (from 200 to 500) yields an approximately 2.4-fold runtime increase across all tested configurations, which is consistent with the theoretical complexity $O(n{\cdot}F{\cdot}S)$ derived in Section 3.6.

Third, even under the most demanding configuration examined (500 orders, 10 factories, and 6 stages), the average runtime remains below 16 s, underscoring the practical viability of I-MOEA/D for large-scale industrial deployment. Taken together, these findings confirm that the proposed algorithm achieves both computational efficiency and scalability, meeting the stringent time constraints typical of real-world production scheduling.

4.5.2. Solution Quality Verification

While the heatmap analysis in Section 4.5.1 confirms computational efficiency, it does not capture solution quality (tardiness and cost). To examine whether I-MOEA/D maintains acceptable solution quality under extreme configurations, a supplementary experiment was conducted on representative configurations: $F \in 3, 5, 10$ (keeping S = 3) and $S \in 2, 4, 6$ (keeping F = 3), using five independent seeds on 200-order instances. Table 8 summarizes the experimental results.

Table 8. Performance of I-MOEA/D under representative factory and stage configurations (5 seeds).

Configuration	F	S	Tardiness (h)	Cost (CNY)	Time (s)
Baseline	3	3	2.28 ± 0.22	125,525 ± 865	5.79 ± 0.14
More Factories	5	3	3.32 ± 0.33	127,570 ± 1204	5.26 ± 0.13
Extreme Factories	10	3	3.09 ± 0.17	138,854 ± 1037	5.43 ± 0.09
Fewer Stages	3	2	1.95 ± 0.16	84,659 ± 1529	5.48 ± 0.05
More Stages	3	4	2.78 ± 0.41	164,881 ± 1122	6.35 ± 0.15
Extreme Stages	3	6	4.49 ± 0.58	246,027 ± 2071	6.94 ± 0.09

As F increases from 3 to 10, tardiness remains within a narrow range (2.28–3.32 h), while total cost rises from 125,525 CNY to 138,854 CNY (+10.6%). This cost increase is primarily attributed to the fixed setup cost: activating 10 factories incurs 20,000 CNY in setup costs versus 6000 CNY for 3 factories—a 14,000 CNY difference that accounts for nearly all of the observed cost increase. When setup costs are excluded, the production and transportation costs actually decrease from 119,525 CNY (3 factories) to 118,854 CNY (10 factories), confirming that additional factories provide genuine operational efficiency gains. Computation time remains stable (5.26–5.79 s) across all factory configurations, indicating that the runtime is not sensitive to F.

As S increases from 2 to 6, both tardiness and cost rise sharply: tardiness increases from 1.95 h to 4.49 h (+130%), and cost increases from 84,659 CNY to 246,027 CNY (+190%). This is a direct consequence of longer production lead times: with 6 stages, the average processing time per order doubles from approximately 4 h (2 stages × 2 h) to 12 h (6 stages × 2 h), leading to more due-date violations and higher production costs. Computation time increases only moderately from 5.48 s to 6.94 s (+27%), confirming that the decoding overhead scales sub-linearly with S, consistent with the theoretical complexity $O(n \cdot F \cdot S)$.

These results confirm that I-MOEA/D maintains reasonable solution quality across a wide range of practical manufacturing configurations, with the stage count having a more substantial effect than the factory count due to its direct impact on processing times. The algorithm's runtime remains well within acceptable limits for real-world decision support, even under extreme configurations. The findings further validate that the conclusions drawn from the main experiments (where $F = 3$ and $S = 3$) are not specific to a particular configuration, but generalize to diverse distributed manufacturing environments.

5. Conclusions

This paper addresses the integrated dynamic production and distribution scheduling problem in a distributed hybrid flow shop environment. A mathematical model is established, and an improved decomposition-based multi-objective evolutionary algorithm, I-MOEA/D, is proposed. Through emergency window management, a 2-opt local search enhancement for vehicle routing, and multiple computational acceleration strategies, the algorithm efficiently converges to high-quality Pareto-optimal solutions. Extensive multi-scale simulation experiments validate its significant advantages over representative existing algorithms in terms of solution quality, robustness, computational efficiency, and schedule stability.

From an academic perspective, this work proposes an integrated model that unifies dynamic order arrivals with distributed production–distribution coordinated optimization, creating a new problem formulation for dynamic scheduling. The emergency window management strategy and 2-opt local search serve as efficient algorithmic components for large-scale NP-hard problems, while the immediate commitment mechanism theoretically eliminates schedule nervousness and provides a rigorous mathematical foundation for stability.

From a practical perspective, extensive experiments demonstrate that I-MOEA/D can generate high-quality schedules within seconds, meeting the real-time decision-making requirements of highly dynamic e-commerce scenarios such as promotional events. The algorithm achieves substantial reductions in both average order tardiness and operational cost while incurring only a marginal loss in on-time delivery rate, reflecting service fairness. The zero-nervousness property avoids shop-floor disruptions caused by frequent plan revisions, ensuring robust engineering feasibility for real-world deployment.

Despite the promising results, several limitations of this study warrant acknowledgment. First, the mathematical model relies on several simplifying assumptions, including homogeneous vehicles, constant travel speeds, and the exclusion of machine breakdowns, traffic variations, and setup times. While these assumptions are commonly adopted in the literature to render the formulation tractable, they may constrain the direct applicability of the proposed approach to highly dynamic real-world environments. Future work could systematically relax these assumptions by incorporating stochastic travel times, heterogeneous vehicle fleets, and machine availability uncertainties. Second, although the scalability analysis in Section 4.5 has validated the algorithm on instances with up to 500 orders, it has only been tested on instances with up to 10 factories and 6 stages. The performance on substantially larger systems—for instance, with 1000 orders and 50 factories—remains to be empirically examined, though the linear time complexity $O(n{\cdot}F{\cdot}S)$ suggests that such extensions would remain computationally viable. We plan to conduct larger-scale validations in future work. Third, the immediate commitment policy, while deliberately designed to ensure zero system nervousness, inherently limits adaptability: orders assigned in earlier windows cannot be reassigned in response to later urgent arrivals. This stability-flexibility trade-off is intentional and well justified in the context of e-commerce fulfillment, where shop-floor disruptions are costly. Nevertheless, scenarios that demand higher adaptability—for example, those with highly uncertain order arrivals—may benefit from a partial relaxation of this policy, such as allowing adjustments for orders with loose due dates. Fourth, all instances used in this study are synthetically generated. Although the generation process is carefully designed to approximate realistic industrial conditions, the absence of validation on publicly available real-world datasets constrains the direct generalizability of our quantitative findings, as standardized benchmarks for this problem class are not yet available. These limitations—concerning the incorporation of more realistic operational conditions, large-scale system validation, and adaptive commitment strategies—constitute our primary agenda for future investigation.

Future research can be pursued along the following directions: incorporating more complex manufacturing configurations such as job shops and open shops in the production stage; introducing practical constraints like multi-type vehicles and time windows in the distribution stage; and at the algorithmic level, further exploring the deep integration of machine learning techniques such as Q-learning with swarm intelligence algorithms to enable adaptive adjustment of operator usage strategies.

Author Contributions: Conceptualization, M.L.; methodology, Z.J.; software, Z.J.; validation, Z.J.; formal analysis, Z.J.; investigation, Z.J.; resources, M.L.; data curation, Z.J.; writing—original draft preparation, Z.J.; writing—review and editing, M.L.; visualization, Z.J.; supervision, M.L.; project administration, M.L.; funding acquisition, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

2-opt	Two-opt (local search operator)
BCMA	Bi-Population Cooperative Memetic Algorithm
CPU	Central Processing Unit
DDHFSVRP	Dynamic Distributed Hybrid Flow Shop Scheduling and Vehicle Routing Problem
DHFS	Distributed Hybrid Flow Shop
DHFSP	Distributed Hybrid Flow Shop Scheduling Problem
DHPFSP-SDST	Distributed Heterogeneous Permutation Flow Shop Scheduling Problem with Sequence-Dependent Setup Times
DRL	Deep Reinforcement Learning
EAM	Earliest Available Machine
EDD	Earliest Due Date
EMO	Evolutionary Multi-Objective Optimization
GA-PSeOA	Hybrid Genetic and Penguin Search Optimization Algorithm
IGD	Inverted Generational Distance
I-MOEA/D	Improved Multi-Objective Evolutionary Algorithm based on Decomposition
IPDSP	Integrated Production and Distribution Scheduling Problem
LMPSO	Level-based Multi-Objective Particle Swarm Optimization
MDVRP	Multi-Depot Vehicle Routing Problem
MILP	Mixed-Integer Linear Programming
MOALNS	Multi-Objective Adaptive Large Neighborhood Search
MOEA/D	Multi-Objective Evolutionary Algorithm based on Decomposition
NP-hard	Nondeterministic Polynomial-time hard
NSGA-II	Nondominated Sorting Genetic Algorithm II
RAM	Random Access Memory
RL-HH	Reinforcement Learning-driven Hyper-Heuristic
SBX	Simulated Binary Crossover
SPT	Shortest Processing Time
FCFS	First-Come-First-Served
LPT	Longest Processing Time

References

1. Maddikunta, P.K.R.; Pham, Q.V.; Prabadevi, B.; Deepa, N.; Dev, K.; Gadekallu, T.R.; Ruby, R.; Liyanage, M. Industry 5.0: A survey on enabling technologies and potential applications. *J. Ind. Inf. Integr.* **2022**, *26*, 100257. [CrossRef]
2. Cai, J.; Lei, D.; Wang, J.; Wang, L. A novel shuffled frog-leaping algorithm with reinforcement learning for distributed assembly hybrid flow shop scheduling. *Int. J. Prod. Res.* **2022**, *61*, 1233–1251. [CrossRef]
3. Azad, T.; Rahman, H.F.; Chakrabortty, R.K.; Ryan, M.J. Optimization of integrated production scheduling and vehicle routing problem with batch delivery to multiple customers in supply chain. *Memetic Comput.* **2022**, *14*, 355–376. [CrossRef]
4. Fu, Y.; Li, Z.; Gao, K.; Fathollahi-Fard, A.M.; Zhang, Z. Integrated scheduling of distributed manufacturing with assembly and distribution: State of the art, challenges, and future directions. *Int. J. Prod. Res.* **2025**, *63*, 1–42. [CrossRef]
5. Chen, Z.L. Integrated Production and Outbound Distribution Scheduling: Review and Extensions. *Oper. Res.* **2010**, *58*, 130–148. [CrossRef]
6. Chen, Z.L. Online Integrated Production and Distribution Scheduling: Review and Extensions. *Inf. J. Comput.* **2025**, *37*, 360–380. [CrossRef]
7. Liu, T.; Liu, J.; Qiu, J.; Lai, C.; Zhang, Z. Recent advances in assembly flow shop scheduling and its extensions to distributed manufacturing and integrated supply chain. *Comput. Ind. Eng.* **2026**, *212*, 111711. [CrossRef]
8. Mohammadi, S.; Al-e-Hashem, S.M.J.M.; Rekik, Y. An integrated production scheduling and delivery route planning with multi-purpose machines: A case study from a furniture manufacturing company. *Int. J. Prod. Econ.* **2020**, *219*, 347–359. [CrossRef]

9. Luo, Q.; Fan, Q.; Deng, Q.; Guo, X.; Gong, G.; Liu, X. Solving bi-objective integrated scheduling problem of production, inventory and distribution using a modified NSGA-II. *Expert Syst. Appl.* **2023**, *225*, 120074. [CrossRef]
10. Long, J.; Pardalos, P.M.; Li, C. Level-based multi-objective particle swarm optimizer for integrated production scheduling and vehicle routing decision with inventory holding, delivery, and tardiness costs. *Int. J. Prod. Res.* **2021**, *60*, 3319–3338. [CrossRef]
11. Horváth, M. New computational results for integrated production and outbound distribution scheduling problems for a product with a short lifespan. *EURO J. Comput. Optim.* **2024**, *12*, 100095. [CrossRef]
12. Sidki, M.; Tchernev, N.; Féniès, P.; Ren, L.; Elfirdoussi, S. A monolithic batch-centric MILP approach for a real-world integrated production and pipeline distribution scheduling problem. *Comput. Ind. Eng.* **2025**, *203*, 111028. [CrossRef]
13. Kaviyani-Charati, M.; Defersha, F.M. Integrated flexible production scheduling and vehicle routing with MLOR constraints: A hybrid metaheuristic approach. *Comput. Ind. Eng.* **2026**, *214*, 111865. [CrossRef]
14. Azad, T.; Rahman, H.F.; Essam, D.; Chakrabortty, R.K. Optimizing sustainable production scheduling and routing in supply chains. *Comput. Ind. Eng.* **2026**, *212*, 111712. [CrossRef]
15. Erdogan, B.; Dang, Q.V.; Mohammadi, M.; Adan, I. A matheuristic for integrated scheduling of manufacturing assembly, product transport, and material supply. *Eur. J. Oper. Res.* **2026**, *in press*. [CrossRef]
16. Wang, Z.X.; Zhang, B.; Li, J.X.; Meng, L.L.; Liu, L.L.; Li, Y.L. A knowledge-guided evolutionary algorithm for integrated group production and transportation scheduling with inter-cell coordination in cellular manufacturing systems. *Comput. Ind. Eng.* **2026**, *219*, 112214. [CrossRef]
17. Duan, J.; Wang, M.; Zhang, Q.; Qin, J. Distributed shop scheduling: A comprehensive review on classifications, models and algorithms. *Math. Biosci. Eng.* **2023**, *20*, 15265–15308. [CrossRef] [PubMed]
18. Fu, Y.; Hou, Y.; Chen, Z.; Pu, X.; Gao, K.; Sadollah, A. Modelling and scheduling integration of distributed production and distribution problems via black widow optimization. *Swarm Evol. Comput.* **2022**, *68*, 101015. [CrossRef]
19. Shao, W.; Shao, Z.; Pi, D. Modeling and multi-neighborhood iterated greedy algorithm for distributed hybrid flow shop scheduling problem. *Knowl.-Based Syst.* **2020**, *194*, 105527. [CrossRef]
20. Wang, J.; Wang, L. A Bi-Population Cooperative Memetic Algorithm for Distributed Hybrid Flow-Shop Scheduling. *IEEE Trans. Emerg. Top. Comput. Intell.* **2021**, *5*, 947–961. [CrossRef]
21. Liu, F.; Gao, K.; Li, D.; Sadollah, A. Ensemble evolutionary algorithms equipped with Q-learning strategy for solving distributed heterogeneous permutation flowshop scheduling problems considering sequence-dependent setup time. *IET Collab. Intell. Manuf.* **2024**, *6*, e12099. [CrossRef]
22. Tibaldo, A.S.; Montagna, J.M.; Fumero, Y. A new integrated approach for solving batching, production scheduling and delivery problems in single-stage batch environments. *Comput. Chem. Eng.* **2024**, *188*, 108753. [CrossRef]
23. Mzili, T.; Mzili, I.; Riffi, M.E.; Pamucar, D.; Simic, V.; Abualigah, L.; Almohsen, B. Hybrid Genetic and Penguin Search Optimization Algorithm (GA-PSeOA) for Efficient Flow Shop Scheduling Solutions. *Facta Univ. Ser. Mech. Eng.* **2024**, *22*, 077–100. [CrossRef]
24. Ouelhadj, D.; Petrovic, S. A survey of dynamic scheduling in manufacturing systems. *J. Sched.* **2009**, *12*, 417–431. [CrossRef]
25. Ding, J.; Chen, M.; Wang, T.; Zhou, J.; Fu, X.; Li, K. A Survey of AI-enabled Dynamic Manufacturing Scheduling: From Directed Heuristics to Autonomous Learning. *ACM Comput. Surv.* **2023**, *55*, 1–36. [CrossRef]
26. Yang, S.; Wang, J.; Xin, L.; Xu, Z. Real-time and concurrent optimization of scheduling and reconfiguration for dynamic reconfigurable flow shop using deep reinforcement learning. *CIRP J. Manuf. Sci. Technol.* **2023**, *40*, 243–252. [CrossRef]
27. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [CrossRef]
28. Zhang, Q.; Li, H. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Trans. Evol. Comput.* **2007**, *11*, 712–731. [CrossRef]
29. Lárraga, G.; Miettinen, K. Survey of Interactive Evolutionary Decomposition-Based Multiobjective Optimization Methods. *Evol. Comput.* **2026**, *in press*. [CrossRef] [PubMed]
30. Wang, Z.; He, M.; Wu, J.; Chen, H.; Cao, Y. An improved MOEA/D for low-carbon many-objective flexible job shop scheduling problem. *Comput. Ind. Eng.* **2024**, *188*, 109926. [CrossRef]
31. Chen, N.; Kang, W.; Kang, N.; Qi, Y.; Hu, H. Order processing task allocation and scheduling for E-order fulfilment. *Int. J. Prod. Res.* **2022**, *60*, 4253–4267. [CrossRef]
32. Risberg, A. A systematic literature review on e-commerce logistics: Towards an e-commerce and omni-channel decision framework. *Int. Rev. Retail Distrib. Consum. Res.* **2022**, *33*, 67–91. [CrossRef]
33. Li, R.; Wang, L.; Gong, W.; Chen, J.; Pan, Z.; Wu, Y.; Yu, Y. Evolutionary computation and reinforcement learning integrated algorithm for distributed heterogeneous flowshop scheduling. *Eng. Appl. Artif. Intell.* **2024**, *135*, 108775. [CrossRef]
34. Ishibuchi, H.; Masuda, H.; Tanigaki, Y.; Nojima, Y. Modified Distance Calculation in Generational Distance and Inverted Generational Distance. In *Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2015; pp. 110–125. [CrossRef]

35. Zitzler, E.; Thiele, L.; Laumanns, M.; Fonseca, C.M.; da Fonseca, V.G. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Trans. Evol. Comput.* **2003**, *7*, 117–132. [CrossRef]
36. Rodriguez-Fernandez, A.E.; Wang, H.; Schütze, O. Reference Set Generator: A Method for Pareto Front Approximation and Reference Set Generation. *Mathematics* **2025**, *13*, 1626. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

MDPI Books Editorial Office
E-mail: books@mdpi.com
www.mdpi.com/books

Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Topic Editors. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editors and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-8646-3