computers

Special Issue Reprint

Transformative Approaches in Education

Harnessing AI, Augmented Reality, and Virtual Reality for Innovative Teaching and Learning

Edited by
Stamatios Papadakis

mdpi.com/journal/computers

MDPI

Transformative Approaches in Education: Harnessing AI, Augmented Reality, and Virtual Reality for Innovative Teaching and Learning

Transformative Approaches in Education: Harnessing AI, Augmented Reality, and Virtual Reality for Innovative Teaching and Learning

Guest Editor

Stamatios Papadakis

MDPI

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

Guest Editor
Stamatios Papadakis
Department of Preschool
Education
University of Crete
Crete
Greece

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Computers* (ISSN 2073-431X), freely accessible at: https://www.mdpi.com/journal/computers/special_issues/0X94314MP3.

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 978-3-7258-7224-4 (Hbk)
ISBN 978-3-7258-7225-1 (PDF)
https://doi.org/10.3390/books978-3-7258-7225-1

© 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 Editor . . . . . **vii**

Stamatios Papadakis
Editorial "Transformative Approaches in Education: Harnessing AI, Augmented Reality, and Virtual Reality for Innovative Teaching and Learning"
Reprinted from: *Computers* **2026**, *15*, 72, https://doi.org/10.3390/computers15020072 . . . . . **1**

Konstantinos Kotsidis, Georgios Chionas and Panagiotes Anastasiades
Exploring Greek Upper Primary School Students' Perceptions of Artificial Intelligence: A Qualitative Study Across Cognitive, Emotional, Behavioral, and Ethical Dimensions
Reprinted from: *Computers* **2026**, *15*, 14, https://doi.org/10.3390/computers15010014 . . . . . **4**

Daniel-Florin Dosaru, Alexandru-Corneliu Olteanu and Nicolae Țăpuș
Optimizing Cloudlets for Faster Feedback in LLM-Based Code-Evaluation Systems
Reprinted from: *Computers* **2025**, *14*, 557, https://doi.org/10.3390/computers14120557 . . . . . **19**

Islam A. Alshafei, Samah Mohammed AlDweik, Mahmoud ali Hassouneh, Hanan AbuKarki, Abdellatif A. Jarrar and Qusai S. Mansour
Digital Transformation in Design Education: Exploring the Challenges and Opportunities in Jordanian Higher Education
Reprinted from: *Computers* **2025**, *14*, 535, https://doi.org/10.3390/computers14120535 . . . . . **33**

Sehar Shahzad Farooq, Hameedur Rahman, Samiya Abdul Wahid, Muhammad Alyan Ansari, Saira Abdul Wahid and Hosu Lee
Curiosity-Driven Exploration in Reinforcement Learning: An Adaptive Self-Supervised Learning Approach for Playing Action Games
Reprinted from: *Computers* **2025**, *14*, 434, https://doi.org/10.3390/computers14100434 . . . . . **70**

Safira Adine Kinari, Nobuo Funabiki, Soe Thandar Aung and Htoo Htoo Sandi Kyaw
A Guided Self-Study Platform of Integrating Documentation, Code, Visual Output, and Exercise for Flutter Cross-Platform Mobile Programming
Reprinted from: *Computers* **2025**, *14*, 417, https://doi.org/10.3390/computers14100417 . . . . . **95**

Athanasios Angeioplastis, John Aliprantis, Markos Konstantakis, Dimitrios Varsamis and Alkiviadis Tsimpiris
The Learning Style Decoder: FSLSM-Guided Behavior Mapping Meets Deep Neural Prediction in LMS Settings
Reprinted from: *Computers* **2025**, *14*, 377, https://doi.org/10.3390/computers14090377 . . . . . **122**

Andrei-Bogdan Stănescu, Sébastien Travadel and Răzvan-Victor Rughiniș
Virtual Reality for Hydrodynamics: Evaluating an Original Physics-Based Submarine Simulator Through User Engagement
Reprinted from: *Computers* **2025**, *14*, 348, https://doi.org/10.3390/computers14090348 . . . . . **151**

Zoltán Szűts, Tünde Lengyelné Molnár, Réka Racskó, Geoffrey Vaughan, Szabolcs Ceglédi and Dalma Lilla Dominek
Examining the Flow Dynamics of Artificial Intelligence in Real-Time Classroom Applications
Reprinted from: *Computers* **2025**, *14*, 275, https://doi.org/10.3390/computers14070275 . . . . . **182**

Michele Domenico Todino, Eliza Pitri, Argyro Fella, Antonia Michaelidou, Lucia Campitiello, Francesca Placanica, et al.
Bridging Tradition and Innovation: Transformative Educational Practices in Museums with AI and VR
Reprinted from: *Computers* **2025**, *14*, 257, https://doi.org/10.3390/computers14070257 . . . . . **199**

Rawan N. Abulail, Omar N. Badran, Mohammad A. Shkoukani and Fandi Omeish
Exploring the Factors Influencing AI Adoption Intentions in Higher Education: An Integrated Model of DOI, TOE, and TAM
Reprinted from: *Computers* **2025**, *14*, 230, https://doi.org/10.3390/computers14060230 . . . . . **212**

Phillip Olla, Ashlee Barnes, Lauren Elliott, Mustafa Abumeeiz, Venus Olla and Joseph Tan
Deploying a Mental Health Chatbot in Higher Education: The Development and Evaluation of Luna, an AI-Based Mental Health Support System
Reprinted from: *Computers* **2025**, *14*, 227, https://doi.org/10.3390/computers14060227 . . . . . **248**

Alejandro De la Hoz Serrano, Andrés Álvarez-Murillo, Eladio José Fernández Torrado, Miguel Ángel González Maestre and Lina Viviana Melo Niño
A Case Study of Computational Thinking Analysis Using SOLO Taxonomy in Scientific–Mathematical Learning
Reprinted from: *Computers* **2025**, *14*, 192, https://doi.org/10.3390/computers14050192 . . . . . **268**

Georgia Iatraki, Tassos A. Mikropoulos, Panos Mallidis-Malessas and Carolina Santos
Leveraging Technology to Break Barriers in Public Health for Students with Intellectual Disabilities
Reprinted from: *Computers* **2025**, *14*, 169, https://doi.org/10.3390/computers14050169 . . . . . **285**

Maria Perifanou and Anastasios A. Economides
Students Collaboratively Prompting ChatGPT
Reprinted from: *Computers* **2025**, *14*, 156, https://doi.org/10.3390/computers14050156 . . . . . **302**

Spyridon Aravantinos, Konstantinos Lavidas, Vassilis Komis, Thanassis Karalis and Stamatios Papadakis
Artificial Intelligence in K-12 Education: A Systematic Review of Teachers' Professional Development Needs for AI Integration
Reprinted from: *Computers* **2026**, *15*, 49, https://doi.org/10.3390/computers15010049 . . . . . **327**

Chrysovalantis Kefalis, Constantine Skordoulis and Athanasios Drigas
A Systematic Review of Mind Maps, STEM Education, Algorithmic and Procedural Learning
Reprinted from: *Computers* **2025**, *14*, 204, https://doi.org/10.3390/computers14060204 . . . . . **357**

About the Editor

Stamatios Papadakis

Stamatios Papadakis is affiliated with the Department of Preschool Education at the University of Crete, Greece. Since 2024, he has served as an Assistant Professor of Educational Technology with an emphasis on mobile learning; prior to this (from 2017 to 2024), he worked as an Adjunct Lecturer in Teaching Informatics in the same department. His scientific and research interests focus on mobile learning, particularly the application of smart mobile devices and their accompanying mobile applications (apps) in preschool and primary education, with a specific focus on the development of computational thinking and students' understanding of numbers. He is currently conducting research on how a STEM learning approach can influence learning through a context-aware mobile learning environment in preschool classrooms, as well as its impact on preschoolers' learning outcomes. His achievements include long-term engagement in educational technology research and teaching, with in-depth exploration of mobile learning's application in early childhood education, providing valuable insights into the integration of technology and preschool education.

computers

MDPI

Editorial

Editorial "Transformative Approaches in Education: Harnessing AI, Augmented Reality, and Virtual Reality for Innovative Teaching and Learning"

Stamatios Papadakis

Department of Preschool Education, University of Crete, 74100 Rethimno, Greece; stpapadakis@uoc.gr

1. Introduction

When we first conceptualized this Special Issue, the educational community was arguably in a state of reaction—reacting to the sudden accessibility of generative AI, the maturing of immersive hardware, and the urgent post-pandemic need for digital resilience. However, the sixteen articles collected in this Special Issue, "Transformative Approaches in Education: Harnessing AI, Augmented Reality, and Virtual Reality for Innovative Teaching and Learning", demonstrate that we have moved beyond reaction to a phase of rigorous, evidence-based integration.

This Special Issue does not merely celebrate new tools; it interrogates them. From the neural architecture of reinforcement learning algorithms to the emotional perceptions of sixth-graders in Greece, the research presented here offers a panoramic view of a field that is rapidly finding its footing. We are witnessing a shift; technology is no longer an external "add-on" but a fundamental structural element of modern pedagogy.

2. The Human Element: Perceptions, Emotions, and Well-being

A defining strength of this Special Issue is its refusal to lose sight of the human being behind the screen. Contribution 1, by Kotsidis et al., opens up a fascinating window into the minds of "digital natives", revealing that Greek primary students have a "hybrid" understanding of AI—perceiving it simultaneously as a helpful robotic assistant and a potential threat. This nuanced ambivalence suggests that AI literacy must address emotional regulation as much as technical skill.

This human-centric focus extends to higher education, where the dynamic between student and machine is evolving into a partnership. Contribution 2, by Perifanou and Economides, documents the emergence of "collaborative prompting", where students effectively treat ChatGPT version GPT-3.5 as a teammate rather than a search engine. Contribution 3, by Szűts et al., deepens this analysis by applying Flow Theory, suggesting that when AI is integrated seamlessly, it can heighten immersion rather than distract from it.

However, the integration of these tools must also safeguard student welfare. Contribution 4, by Olla et al., presents "Luna", a mental health chatbot, demonstrating that AI can serve as a scalable, immediate first line of defense for student well-being, provided that strict ethical guardrails are maintained.

3. Infrastructures of Adoption and Inclusion

Innovation cannot exist in a vacuum; it requires institutional readiness. Contribution 5, by Abulail et al., utilizes a robust integrated model (DOI, TOE, and TAM) to identify

friction points in higher education, highlighting that despite high perceived utility, institutional infrastructure remains a significant bottleneck. This is echoed in Contribution 6, by Alshafei et al., who survey design education in Jordan. Their findings reveal a "digital divide" within curricula themselves: while traditional CAD tools are ubiquitous, transformative technologies such as AI and VR remain underutilized due to steep learning curves and costs.

Crucially, this Special Issue underscores the fact that technology must serve equity. Contribution 7, by Iatraki et al., provides compelling evidence that Digital Learning Objects (DLOs) can successfully break down barriers in public health education for students with intellectual disabilities. Similarly, in Contribution 8, Aravantinos et al. argue through a systematic review that teacher professional development must pivot from purely technical training to fostering pedagogical reasoning and inclusive practices.

4. The Engineering of Adaptive Learning

Behind the user interface lies the sophisticated engineering that makes adaptive learning possible. Several contributions in this Special Issue push the technical boundaries of what educational systems can achieve. In Contribution 9, Dosaru et al. tackle the scalability crisis in coding education with "LambdaChecker", optimizing cloudlet resource allocation to handle mass-scale automated code evaluation without performance degradation.

On the algorithmic frontier, in Contribution 10, Farooq et al. explore curiosity-driven exploration in reinforcement learning. By designing agents that are intrinsically motivated by novelty, their work offers a blueprint for future adaptive tutoring systems that can "learn to teach" by exploring student needs dynamically. This aligns with Contribution 11, by Angeioplastis et al., whose "Learning Style Decoder" utilizes deep neural networks to predict student preferences from LMS behavior, moving us closer to truly personalized education.

5. Cognitive Frameworks and Immersive Realities

Finally, we see how immersive tools are reshaping the cognitive architecture of learning. Contribution 12, by Stănescu et al., demonstrates that high-fidelity VR is not just a gimmick but a necessity for teaching counter-intuitive physics concepts such as hydrodynamics. In the cultural sector, Contribution 13, by Todino et al., shows how VR and AI can revitalize museum education, transforming static folk art into living experiences.

However, traditional cognitive structures remain vital. Both Contribution 14 by De la Hoz Serrano et al. and Contribution 15 by Kefalis et al. reinforce that whether we use educational robotics or digital mind maps, the development of computational thinking relies on structured pedagogical taxonomies (such as SOLO) to organize complex information. Contribution 16, by Kinari et al., bridges this gap between structure and freedom with a guided self-study platform for mobile programming, proving that documentation and visual output must be integrated to prevent "copy–paste" learning.

6. Conclusions

The diversity of these sixteen contributions—spanning from the algorithmic "curiosity" of machines to the very human curiosity of children—confirms that we are in a transformative era. We are no longer asking *if* AI and VR have a place in education; we are defining the specific, evidence-based methodologies for *how* they should be used.

Due to the overwhelming success and high citation potential of these articles, I am pleased to announce the launch of a **Second Edition** of this Special Issue. We invite the academic community to continue this vital conversation.

https://doi.org/10.3390/computers15020072

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflicts of interest.

List of Contributions:

1. Kotsidis, K.; Chionas, G.; Anastasiades, P. Exploring Greek Upper Primary School Students' Perceptions of Artificial Intelligence: A Qualitative Study Across Cognitive, Emotional, Behavioral, and Ethical Dimensions. *Computers* **2026**, *15*, 14.
2. Perifanou, M.; Economides, A.A. Students Collaboratively Prompting ChatGPT. *Computers* **2025**, *14*, 156.
3. Szűts, Z.; Lengyelné Molnár, T.; Racskó, R.; Vaughan, G.; Ceglédi, S.; Dominek, D.L. Examining the Flow Dynamics of Artificial Intelligence in Real-Time Classroom Applications. *Computers* **2025**, *14*, 275.
4. Olla, P.; Barnes, A.; Elliott, L.; Abumeeiz, M.; Olla, V.; Tan, J. Deploying a Mental Health Chatbot in Higher Education: The Development and Evaluation of Luna, an AI-Based Mental Health Support System. *Computers* **2025**, *14*, 227.
5. Abulail, R.N.; Badran, O.N.; Shkoukani, M.A.; Omeish, F. Exploring the Factors Influencing AI Adoption Intentions in Higher Education: An Integrated Model of DOI, TOE, and TAM. *Computers* **2025**, *14*, 230.
6. Alshafei, I.A.; AlDweik, S.M.; Hassouneh, M.A.; AbuKarki, H.; Jarrar, A.A.; Mansour, Q.S. Digital Transformation in Design Education: Exploring the Challenges and Opportunities in Jordanian Higher Education. *Computers* **2025**, *14*, 535.
7. Iatraki, G.; Mikropoulos, T.A.; Mallidis-Malessas, P.; Santos, C. Leveraging Technology to Break Barriers in Public Health for Students with Intellectual Disabilities. *Computers* **2025**, *14*, 169.
8. Aravantinos, S.; Lavidas, K.; Komis, V.; Karalis, T.; Papadakis, S. Artificial Intelligence in K-12 Education: A Systematic Review of Teachers' Professional Development Needs for AI Integration. *Computers* **2026**, *15*, 49.
9. Dosaru, D.-F.; Olteanu, A.-C.; Țăpuș, N. Optimizing Cloudlets for Faster Feedback in LLM-Based Code-Evaluation Systems. *Computers* **2025**, *14*, 557.
10. Farooq, S.S.; Rahman, H.; Abdul Wahid, S.; Alyan Ansari, M.; Abdul Wahid, S.; Lee, H. Curiosity-Driven Exploration in Reinforcement Learning: An Adaptive Self-Supervised Learning Approach for Playing Action Games. *Computers* **2025**, *14*, 434.
11. Angeioplastis, A.; Aliprantis, J.; Konstantakis, M.; Varsamis, D.; Tsimpiris, A. The Learning Style Decoder: FSLSM-Guided Behavior Mapping Meets Deep Neural Prediction in LMS Settings. *Computers* **2025**, *14*, 377.
12. Stănescu, A.-B.; Travadel, S.; Rughiniş, R.-V. Virtual Reality for Hydrodynamics: Evaluating an Original Physics-Based Submarine Simulator Through User Engagement. *Computers* **2025**, *14*, 348.
13. Todino, M.D.; Pitri, E.; Fella, A.; Michaelidou, A.; Campitiello, L.; Placanica, F.; Di Tore, S.; Sibilio, M. Bridging Tradition and Innovation: Transformative Educational Practices in Museums with AI and VR. *Computers* **2025**, *14*, 257.
14. De la Hoz Serrano, A.; Álvarez-Murillo, A.; Fernández Torrado, E.J.; González Maestre, M.Á.; Melo Niño, L.V. A Case Study of Computational Thinking Analysis Using SOLO Taxonomy in Scientific-Mathematical Learning. *Computers* **2025**, *14*, 192.
15. Kefalis, C.; Skordoulis, C.; Drigas, A. A Systematic Review of Mind Maps, STEM Education, Algorithmic and Procedural Learning. *Computers* **2025**, *14*, 204.
16. Kinari, S.A.; Funabiki, N.; Aung, S.T.; Kyaw, H.H.S. A Guided Self-Study Platform of Integrating Documentation, Code, Visual Output, and Exercise for Flutter Cross-Platform Mobile Programming. *Computers* **2025**, *14*, 417.

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.

https://doi.org/10.3390/computers15020072

computers

MDPI

Article

Exploring Greek Upper Primary School Students' Perceptions of Artificial Intelligence: A Qualitative Study Across Cognitive, Emotional, Behavioral, and Ethical Dimensions

Konstantinos Kotsidis [1,*], Georgios Chionas [2] and Panagiotes Anastasiades [1]

[1] Laboratory of Advanced Learning Technologies in Lifelong and Distance Learning (EDIVEA), University of Crete, 74100 Rethymnon, Greece; anastasiadis@uoc.gr
[2] 1st Primary School of Elliniko, Iasonidou 52, 16777 Elliniko, Greece; gchionas@sch.gr
* Correspondence: kkotsidis@edc.uoc.gr

Abstract

This study investigates the perceptions of Greek sixth-grade students regarding Artificial Intelligence (AI). Understanding students' pre-instructional conceptions is essential for developing targeted interventions that build on existing knowledge rather than assuming conceptual deficits. A qualitative design was employed with 229 students from seven elementary schools in Athens, Greece. Data were collected through open-ended questions and word association tasks, then analyzed using Walan's AI perceptions framework as an integrated set of analytical lenses (cognitive, affective, behavioral/use, and ethical considerations). Findings revealed that students hold multifaceted conceptions of AI. Cognitively, they described AI as robots, computational systems, software tools, and autonomous learning programs. Affectively, they expressed ambivalence, balancing appreciation of AI's usefulness with concerns over potential risks. Behaviorally, they identified interactive question–answer functions, creative applications, and everyday assistance roles. Ethically, students raised issues of responsible use, societal implications, and human–AI relationships. This study contributes to international research, highlighting that primary students' understandings of AI are more nuanced than is sometimes assumed, and offer empirical insights for designing culturally responsive, ethically informed AI literacy curricula.

Keywords: AI literacy; artificial intelligence; greek students; primary education; qualitative research; student perceptions

1. Introduction

As Artificial Intelligence (AI) systems become increasingly embedded in everyday technologies, children are exposed to them from a young age—often through educational platforms, voice assistants, or algorithm-driven apps—before any formal instruction occurs [1]. Such early, pre-instructional experiences shape intuitive understandings [2], influencing how students later engage with AI-related learning. Understanding students' conceptions of AI is thus critical for designing meaningful AI literacy education.

Research has shown that, even at the primary level, children form complex representations of AI, often combining factual knowledge with anthropomorphic, fictionalized, or emotionally charged associations [3,4]. Building on these findings, recent studies have explored this phenomenon across diverse contexts and age groups, revealing hybrid views of AI that combine technical understanding with anthropomorphic elements and emotional evaluations [4–6]. Students often associate AI with both support and threat [2,7],

yet their frequent use of AI tools does not necessarily translate into a coherent conceptual understanding of how these systems function.

However, despite these insights, existing research rarely examines children's perceptions across cognitive, emotional, behavioral, and ethical dimensions in an integrated way. To address this gap, Walan [1] proposed an integrated framework for analyzing primary students' AI perceptions, incorporating these four dimensions. Grounded in Mitcham's [8] philosophy of technology and expanded by Ankiewicz [9], the framework offers a structured yet flexible way to interpret children's multifaceted understandings of AI. This study responds to that gap by applying Walan's framework to investigate Greek sixth-grade students' pre-instructional conceptions of AI.

2. Conceptual Framework

This study is situated within the growing field of human-centered Artificial Intelligence (AI) literacy, which conceptualizes AI not merely as a technical artifact to be mastered, but as a socio-technical phenomenon embedded in learners' cognitive representations, emotional attitudes, everyday practices, and emerging ethical stances from early school age.

Contemporary AI literacy research emphasizes the need to move beyond narrow computational or coding-oriented approaches and toward holistic pedagogical frameworks integrating epistemic, affective, behavioral, and moral dimensions of human engagement with intelligent technologies [10]. Nevertheless, much of the existing literature remains fragmented, often focusing either on discrete technical knowledge or on patterns of tool use, and therefore rarely applies integrated conceptual lenses in a systematic manner to capture the multidimensional nature of children's meaning-making around AI. This fragmentation limits cross-study comparability and restricts the development of culturally responsive educational interventions.

In response to this need, the present study adopts Walan's [1] integrated framework for analyzing primary students' perceptions of AI. The framework organizes student conceptions across four dimensions—cognitive, affective, behavioral, and ethical—designed to encapsulate how children conceptualize, emotionally evaluate, practically engage with, and morally reflect upon emerging AI technologies. Theoretically grounded in Mitcham's [8] philosophy of technology and its educational elaboration by Ankiewicz [9], the model approaches technological literacy as a dynamic human–technology relationship. This multidimensional approach aligns with contemporary international research that emphasizes the interplay of these dimensions in understanding AI literacy, including recent studies in diverse contexts such as West Africa [11].

In this study, the four dimensions serve as interconnected and flexible analytical lenses (following Blumer's [12] notion of sensitizing concepts), not as rigid or hierarchical categories. This is particularly evident in the role of the ethical dimension, which functions as a permeating evaluative layer that shapes and emerges within cognitive understandings, affective responses, and behavioral narratives. Figure 1 provides a visual representation of this integrated logic, where the ethical dimension is depicted as encompassing the other three, reflecting its cross-cutting nature. The bidirectional arrows among the cognitive, affective, and behavioral dimensions illustrate their reciprocal influence: cognitive understanding shapes emotional responses and practical engagement, while experiences and feelings recursively reshape conceptual models.

The operational definitions of the four dimensions as applied in this study are summarized below:

- Cognitive dimension: Students' conceptual explanations of AI, including representations of AI as machines, robots, computational systems, software platforms, or autonomous learning entities.

https://doi.org/10.3390/computers15010014

- Affective dimension: Emotional evaluations of AI, encompassing expressions of enthusiasm, fear, trust, skepticism, ambivalence, or anxiety regarding its societal role or potential impact.
- Behavioral dimension: Descriptions of actual or imagined interactions with AI systems, including educational use, creative applications, communication, and everyday assistance practices.
- Ethical dimension: Reflections related to responsibility, human agency, social consequences, fairness, misuse, employment displacement, and future human–AI relations.

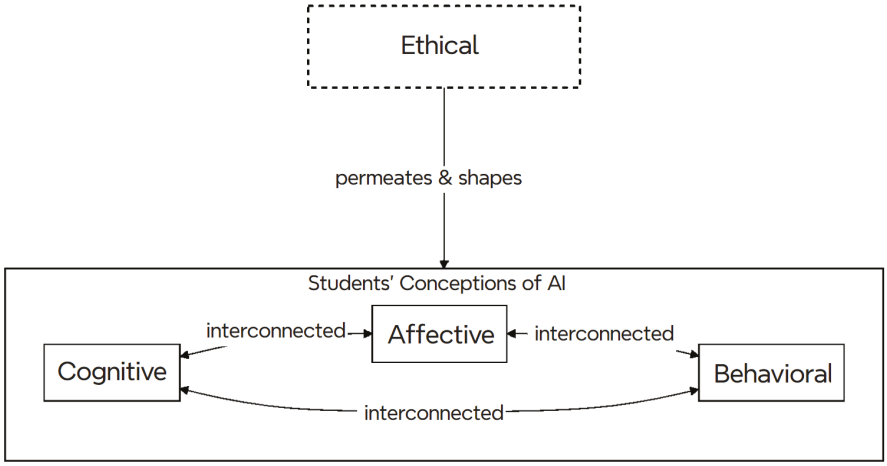


Figure 1. Visual representation of Walan's integrated framework for analyzing primary students' AI perceptions, with the ethical dimension as a permeating layer encompassing the cognitive, affective, and behavioral dimensions.

This study does not treat the application of Walan's framework as an attempt to replicate prior work; instead, it uses the framework to theoretically situate and guide the analysis in relation to prior research on students' perceptions of AI while maintaining a non-evaluative stance toward the framework itself.

Guided by this framework, the present study addresses the following research questions:

RQ1: How do Greek upper primary school students perceive AI in terms of its cognitive dimension?

RQ2: How do Greek upper primary school students perceive AI in terms of its affective dimension?

RQ3: How do Greek upper primary school students perceive AI in terms of its behavioral dimension?

RQ4: How do Greek upper primary school students perceive AI in terms of its ethical dimension?

These questions explore interrelated facets of students' AI perceptions, consistent with the integrated logic of Figure 1.

3. Methodology

3.1. Triangulation and Mixed-Methods Research Design

This study adopted a qualitative inductive design subsequently theoretically informed by Walan's [1] four-dimensional framework. The analysis began inductively through open coding and iterative comparison, allowing patterns to emerge directly from the data. The resulting themes were then interpreted abductively, drawing on the framework as a set of sensitizing lens [12] to connect emerging patterns with existing theory [13]. A detailed account of the analytical phases and procedures is presented in Section 3.5.

https://doi.org/10.3390/computers15010014

3.2. Participants

The study involved 229 sixth-grade students (aged 11–12 years) from seven public elementary schools in the greater Athens metropolitan area. The sample, recruited through convenience sampling, consisted of 124 male (54.1%) and 105 female (45.9%) students, all attending mainstream general education classrooms. Prior to data collection, the study received formal institutional approval from the Greek Ministry of Education (Protocol No. 4830, March 2025) and fully adhered to ethical guidelines for research involving minors. School principals were formally informed in writing about the study's aims and procedures and provided institutional consent. Parents received a written announcement explaining the purpose, voluntary nature, and confidentiality of the research, in line with ethical guidelines for school-based assessment. Verbal consent was obtained from parents, as the activity involved minimal risk and consisted solely of completing an anonymous questionnaire [14].

3.3. Data Collection Procedures

Data collection took place during April 2025, within regular school hours, across seven public elementary schools in Athens. Classroom teachers administered the paper-and-pencil questionnaire using a standardized protocol developed by the research team to ensure consistent procedures across sites. The procedure involved two phases within a single instructional period. In the first phase, students responded to the open-ended question: "What is Artificial Intelligence, in your opinion?". In the second phase, students completed a word-association task: "What words come to mind when you hear 'Artificial Intelligence'? (Write at least 3)".

In line with recent recommendations for child-rights–based and ethically grounded participation in educational research [15], teachers received detailed facilitator guidelines emphasizing procedural neutrality [16]. They were asked to make clear that there were no "right" or "wrong" answers, to encourage students to express their own ideas freely, and to limit any interventions to procedural clarifications only. Teachers were explicitly instructed to avoid providing examples, interpretations, or suggestions that might influence the students' responses.

Participation was entirely voluntary. Students were informed that they could skip any question or withdraw from the activity at any time, and none chose to discontinue their participation. Upon completion, the anonymous questionnaires were collected by the teachers, sealed in envelopes, and securely delivered to the research team for analysis.

3.4. Data Preparation

All data were digitized and organized into two categories: written definitional responses and word associations. All handwritten responses were transcribed verbatim by two independent researchers. To ensure lexical consistency and analytical validity, a systematic data-cleaning process was carried out prior to analysis [17]. This process involved spelling corrections, grammatical normalization, and the consolidation of semantically equivalent terms. For example, responses with similar or identical meanings (e.g., computer, computers, PC, and laptop) were grouped under the unified term Computer, aiming to reduce redundancy and enhance interpretability. The procedure was conducted independently by the first two researchers experienced in qualitative data analysis [18,19] focusing on the preliminary data preparation stage rather than on the analytical categorization itself. Any differences identified during this stage concerned the harmonization of closely related terms, clarification of overlapping expressions, and standardization of similar conceptual references in students' responses. These issues were resolved through

https://doi.org/10.3390/computers15010014

discussion prior to the analytical phase and did not involve disagreements. The cleaned data thus served as the final corpus used for inductive categorization.

3.5. Analytical Framework

This study employed a systematic two-phase analytical process conducted using Microsoft Excel to examine students' AI perceptions while maintaining both empirical openness and theoretical grounding. The use of Microsoft Excel was selected as it provided transparency, ease of reproducibility, and efficient organization, filtering, and triangulation of qualitative data [20]. The subsequent sections describe the two analytical phases through which data were organized and theoretically interpreted. An overview of this qualitative data analysis process is illustrated in Figure 2.

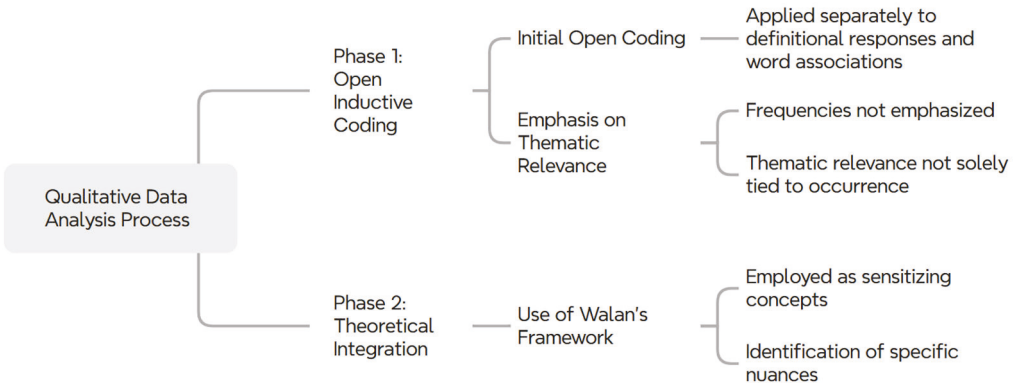


Figure 2. Overview of the Qualitative Data Analysis Process.

3.5.1. Phase 1: Open Inductive Coding

Initial open coding [21] was applied separately to definitional responses and word associations. The former were analyzed to identify conceptual categories and functional meanings, while the latter were grouped into semantic clusters based on similarity. This inductive process involved iterative reading, constant comparison across data segments, and the progressive refinement of codes to capture emerging meanings [22]. Through memoing and reflection, initial codes were reviewed and consolidated into more abstract patterns that informed subsequent thematic integration. Consistent with the study's approach, frequencies were not emphasized, as thematic relevance is not solely tied to occurrence [23].

3.5.2. Phase 2: Theoretical Integration

Emerging themes then were mapped onto Walan's [1] four-dimensional framework which was employed as a set of sensitizing concepts rather than fixed categories [12]. This process enabled both alignment with the existing model and identification of specific nuances [24,25]. In line with the Five-Phase Process of Qualitative Data Analysis proposed by Bingham [22], this stage involved interpreting inductively derived themes through theoretical and conceptual lenses to achieve a transparent and theory-informed integration of findings.

3.6. Final Coding Framework

The final coding framework, summarized in Table 1, illustrates the hierarchical organization of dimensions and categories and served as the analytical foundation for the presentation of results. In line with the use of sensitizing concepts, when student responses exhibited characteristics of multiple categories, coders identified the primary or most prominent conceptual dimension emphasized in the response. For example, a response stating

https://doi.org/10.3390/computers15010014

"the ability of computers or robots to learn and solve problems" (Student 201) was coded under "Autonomous Learning Systems" because the central emphasis was on learning capability, with computers and robots serving as the medium rather than the defining feature. In contrast, a response describing "a computer that thinks and responds like a human" (Student 3) was coded under "Computational Systems" because the focus was on cognitive processing rather than adaptive learning. This interpretive decision-making was guided by the consensus discussions during the sample joint coding phase, ensuring consistent identification of each response's dominant thematic focus without treating numerical intercoder coefficients as a primary quality criterion.

Table 1. Coding Framework with Category Definitions and Key Distinguishing Features.

Dimension	Category	Definition	Key Distinguishing Feature
Cognitive	AI as Robotic Entities	Students describe AI as robots with human-like intelligence, actions, or reasoning.	Primary emphasis on physical embodiment and robotic form.
	AI as Computational Systems	AI understood as advanced computers or machines performing cognitive tasks.	Primary emphasis on computational processing and information handling.
	AI as Software Applications	Students identify AI with specific programs or platforms (e.g., ChatGPT).	Primary emphasis on named applications and specific tools.
	AI as Autonomous Learning Systems	AI perceived as capable of self-learning or adaptation.	Primary emphasis on learning and adaptive capabilities.
Affective	Positive Utility Perceptions	Expressions of enthusiasm and perceived usefulness of AI in learning or daily life.	Predominantly positive emotional valence without balancing concerns.
	Risk and Threat Concerns	Fears of danger, domination, or destructive outcomes associated with AI.	Predominantly negative emotional valence focused on threats and risks.
	Ambivalent Orientations	Simultaneous positive and negative evaluations of AI's role.	Explicit acknowledgment of both positive and negative dimensions.
Behavioral	Question-Response Interactions	Descriptions of interactive communication with AI.	Primary emphasis on interactive dialog as behavioral mode.
	Creative and Educational Applications	AI used or imagined as creative or learning tool.	Primary emphasis on content creation or learning support.
	Everyday Assistance and Support	AI viewed as a practical assistant that simplifies life.	Primary emphasis on practical utility in daily life.
Ethical	Responsible Use Awareness	Recognition of ethical use and human responsibility in interacting with AI.	Primary emphasis on human agency and responsibility in AI use.
	Future Societal Impact	Reflections on AI's effect on jobs, society, and the future of humanity.	Primary emphasis on societal-level consequences beyond individual use.
	Human-AI Relationship Dynamics	Awareness of AI as human creation and acknowledgment of human agency.	Primary emphasis on human authorship and control of AI systems.

3.7. Research Quality

Research quality was addressed through procedures ensuring dependability, credibility, and confirmability [26]. Dependability was supported by standardized research

https://doi.org/10.3390/computers15010014

protocols across schools and joint coding of 22% of responses (n = 50) by two researchers to ensure procedural consistency. This joint coding functioned as a negotiated-agreement process: disagreements were discussed, code definitions and decision rules were refined, and a shared set of analytic guidelines was documented before proceeding. After establishing shared analytical guidelines through discussion and consensus, the second author conducted the complete coding of the remaining dataset following the agreed criteria [27,28]. In line with contemporary debates in qualitative methodology, the study prioritized transparency of the coding frame and decision rules over reporting a single numerical intercoder reliability coefficient [29]. Credibility was enhanced through data (definitions and word-associations) and investigator triangulation, reducing potential bias and supporting authentic representation of students' perspectives. Confirmability was established by maintaining systematic analytic memos and preserving students' original Greek expressions, providing a transparent audit trail. All the aforementioned procedures were consistent with contemporary guidelines for ensuring trustworthiness in qualitative research [29,30].

4. Results

All 229 sixth-grade students who participated in the study provided valid responses. Response quality varied considerably, ranging from single-word answers to detailed multi-sentence explanations, with single-word responses appearing primarily in word association items and more elaborate descriptions observed in definitional responses. All student answers were translated from Greek into English for publication purposes using Anthropic Claude v3.7, with manual unit-by-unit verification (sentences and single-word responses) of the entire dataset to ensure semantic accuracy and preservation of linguistic nuance.

The results are presented according to the four analytical dimensions that guided the study—cognitive, affective, behavioral, and ethical. A summary of all conceptual categories across the four dimensions is presented in Figure 3. This structure reflects the inductive development of categories that were subsequently organized into these theoretical dimensions, allowing for an integrated interpretation of students' conceptual, emotional, behavioral, and moral understandings of AI.

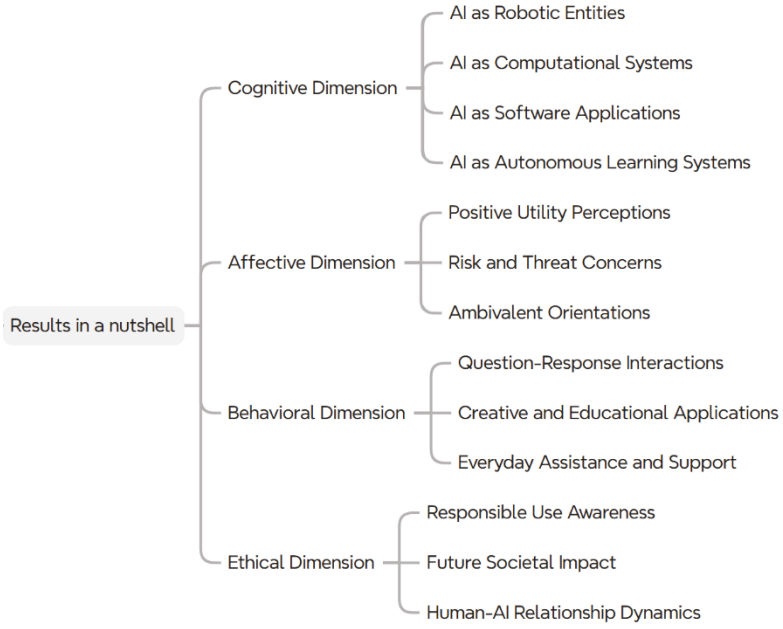


Figure 3. Summary of the conceptual categories derived from students' responses across Walan's [1] four dimensions.

https://doi.org/10.3390/computers15010014

4.1. Cognitive Dimension: Students' Understanding of AI as Technology

Analysis of definitional responses revealed four primary conceptual categories within the Cognitive Dimension, illustrating how Greek primary students conceptualize AI. These categories encompassed perceptions of AI as (1) robotic entities, (2) computational systems, (3) software applications, and (4) autonomous learning systems.

4.1.1. AI as Robotic Entities

Students frequently characterized AI as robotic entities possessing physical embodiment and human-like capabilities. Representative responses included: "I believe that AI is robots that move like humans" (Student 1); "Robots and generally artificial bodies" (Student 94); "A robot made by humans and smarter than humans" (Student 27); "Robots that can do housework and cooking work" (Student 99); "Robots and whatever it can do with its hands" (Student 20).

4.1.2. AI as Computational Systems

Many students conceptualized AI as sophisticated computer systems or machines. Responses included: "A computer that thinks, distinguishes objects and answers questions" (Student 3); "Something like a machine that knows all the answers to our questions" (Student 4); "A gigantic computer that gives us information" (Student 60); "An automatic computer that can do and answer everything" (Student 71).

4.1.3. AI as Software Applications

A substantial number of students referenced specific AI tools and applications reflecting the rapid commercialization of AI tools. Students stated: "An application related to electricity that helps with schoolwork and questions, like ChatGPT" (Student 18); "Various applications like AI which is an artificial friend you can ask various questions" (Student 112); "ChatGPT and Lisari" (Student 155); "All electronic programs like ChatGPT that help us solve daily life problems" (Student 66).

4.1.4. AI as Autonomous Learning Systems

Students described AI as systems capable of independent learning and adaptation, indicating emerging awareness of adaptive capabilities, though not necessarily deep technical understanding. Responses included: "A program that can learn and change appropriately with the information it receives from humans" (Student 21); "The ability of computers or robots to learn, think, and solve problems like human minds" (Student 201); "Programming that has its own thinking" (Student 111); "An artificial brain that is made to help humans in daily life" (Student 134).

Word associations reflected these conceptualizations across all four categories, with frequent references to robots and physical forms, computational systems, branded applications, and adaptive capabilities. Representative associations included: "robot, intelligence, AI" (Student 14), "computer, technology, robot, ChatGPT, AI" (Student 78), "technology, robotics, robot, informatics, programming" (Student 130), and "robot, code, answers, processing, adaptation, algorithm" (Student 189). Notably, word associations revealed stronger brand-specific awareness than definitional responses, with frequent mentions of "Tesla, ChatGPT, Neuralink AI" (Student 44) and "ChatGPT, Microsoft, Siri, Gemini, AI, Apple" (Student 154).

4.2. Affective Dimension: Emotional Responses and Attitudes Toward AI

Student responses included emotional references of varying orientation—positive, negative, and ambivalent—which correspond to Walan's [1] affective dimension and illustrate the range of students' emotional positions toward AI technology.

https://doi.org/10.3390/computers15010014

4.2.1. Positive Utility Perceptions

Many students expressed positive attitudes toward AI, emphasizing its helpful nature in educational and daily contexts. Representative responses included: "Entertaining, it can create any image you want and if you have any question it answers and always helps you" (Student 5); "Like an assistant or friend that can help you with support and give you ideas" (Student 33); "A tool that facilitates our daily lives and helps us with many things" (Student 85); "A great help for humans and for a better life" (Student 108).

4.2.2. Risk and Threat Concerns

Students expressed fears about AI's potential negative impacts. Responses included: "I believe robots should not exist because they are dangerous" (Student 47); "I fear that it might one day conquer the world" (Student 170); "It should not have been created because in the future it might conquer the world" (Student 152); "Something that if it develops too much will cause many disasters" (Student 227).

4.2.3. Ambivalent Orientations

Students demonstrated simultaneous acknowledgment of both positive and negative AI characteristics. Examples included: "AI can do us both good and bad depending on what AI we're talking about" (Student 7); "Sometimes very useful in our days but sometimes can become very bad" (Student 146); "I think it will help partially but it's not ruled out that it could prove fatal" (Student 182).

Word association data triangulated these emotional responses across all three affective orientations. Positive utility associations spanned expressions of helpfulness and enthusiasm ("help, fantasy, technology" [Student 5]; "utility, intelligence, tool, ease" [Student 3]; "smart, assistant, fast answers" [Student 17]). Risk and threat concerns were evident in darker associations ("problems, bad, powerful" [Student 6]; "robot, exterminator, apocalypse" [Student 46]; "total destruction" [Student 140]). Ambivalent orientations combined both perspectives ("domination, help, third world war" [Student 146]; "facilitation, help, fear, worry" [Student 97]).

4.3. Behavioral Dimension: AI Usage and Interaction Patterns

Students described various forms of current and anticipated AI engagement behaviors, reflecting both established patterns from previous research and contemporary interaction modalities.

4.3.1. Question-Response Interactions

Students frequently described AI through interactive question-answering capabilities. Representative responses included: "You can ask it anything and it will answer" (Student 31); "A robot you can ask various things" (Student 12); "Something that answers immediately to whatever you ask" (Student 229); "A system that can answer all your questions" (Student 192).

4.3.2. Creative and Educational Applications

Students demonstrated awareness of AI's creative and educational capabilities. Responses included: "Can create any image you want and help with studies" (Student 5); "Helps with homework and exercises" (Student 132); "Can generate images, solve problems and generally help humans facilitate their daily life" (Student 79); "Programs in robots or computers that help you solve questions or create any image or video" (Student 185).

https://doi.org/10.3390/computers15010014

4.3.3. Everyday Assistance and Support

Students conceptualized AI as providing comprehensive daily life support. Examples included: "A digital assistant that can help you with whatever you want" (Student 119); "A system that facilitates people's lives" (Student 210); "A tool created by humans to help them with daily and normal problems" (Student 226); "An assistant that knows everything and can help you with whatever you need" (Student 198).

Word associations reflected these behavioral dimensions across all three categories. Interactive capabilities were emphasized through direct communication patterns ("AI, ChatGPT, questions, answers" [Student 70]; "fast answers" [Student 17]; "answers to questions" [Student 87]). Creative and educational applications emerged in references to content generation and learning support ("ChatGPT, lessons, teacher" [Student 18]; "homework solutions, help with questions" [Student 155]; "assignments" [Student 185]). Everyday assistance was evident in utility-focused associations ("utility, intelligence, tool, ease" [Student 3]; "convenience, human life" [Student 35]; "assistant, omniscient, helpful" [Student 171]).

4.4. Ethical Dimension: Moral and Social Considerations

Students incorporated ethical considerations into their definitional responses, indicating an awareness of moral aspects associated with AI that extends beyond mere functional understanding.

4.4.1. Responsible Use Awareness

Students demonstrated awareness of responsible AI use. Representative responses included: "However, it is very important that we use it correctly" (Student 3); "We need to be careful what we do because bad people can use AI, for example hackers to harm and deceive people" (Student 30); "Some children use it for cheating but it doesn't always have correct answers" (Student 112).

4.4.2. Future Societal Impact

Students expressed understanding of AI's long-term societal implications. Responses included: "In the future will be very important in life and I believe it will be in our daily routine" (Student 77); "They say that in the future it will replace people's jobs" (Student 85); "A program that in the coming years will dominate the whole world" (Student 178); "The future of humanity and a great ally for now" (Student 145).

4.4.3. Human-AI Relationship Dynamics

Students demonstrated awareness of the constructed nature of AI and human agency in its development. Examples included: "We have created and programmed them" (Student 41); "A creation of humans that helps them in their work and generally in their life" (Student 169); "The intelligence of machines created by humans to serve humans themselves" (Student 206); "A human creation that helps people when they need help sometimes" (Student 217).

Ethical concerns appeared in word associations, predominantly through future societal impact references ("domination, help, third world war" [Student 138]; "world destruction, help, world domination" [Student 71]; "future, robot, destruction, technology" [Student 182]; "robot, future, world upgrade, future destruction" [Student 125]. Human-AI relationship dynamics emerged in creation-oriented associations ("programming, experiments, technology" [Student 169]; "human, intelligence, robot" [Student 19], supporting the spontaneous ethical reasoning identified in definitional responses.

https://doi.org/10.3390/computers15010014

5. Discussion and Implications

The purpose of this study was to explore how Greek sixth-grade students conceptualize AI before receiving any formal instruction, with the broader aim of informing age-appropriate approaches to AI literacy in primary education. Data were collected from written definitional responses and word associations provided by 229 students across seven public schools in Athens. The material was analyzed qualitatively through a two-phase process of open inductive coding followed by theoretical integration using Microsoft Excel. Walan's [1] four-dimensional framework—cognitive, affective, behavioral, and ethical—served as an interpretive lens to organize and interpret the emergent categories. The following discussion interprets the findings with reference to each research question and compares them with previous international studies, aiming to clarify which aspects of students' perceptions are shared across contexts and which appear specific to the Greek sample.

Regarding the cognitive dimension, students described AI in terms that combined human-like cognition with computational processes. They viewed AI as capable of thinking, learning, and adapting, and often connected these ideas with specific technologies such as ChatGPT, Tesla, or Neuralink. This pattern aligns with findings from Swedish and Turkish students, who also described AI as "brain-like" or autonomous [1,6]. In contrast with earlier work where references were largely generic [4], these responses included direct naming of contemporary systems, showing that recent generative AI tools have entered students' linguistic repertoire. Students' emphasis on functions such as answering, creating, and problem solving reflects functional patterns of understanding that have also been described in other European studies [2,4], where AI was similarly portrayed as both computational and adaptive. Anthropomorphic expressions such as "robots that move like humans" or "an artificial brain" appeared frequently, corresponding to Swedish and Turkish students' depictions of AI as embodied or human-like [1,6]. Dutch students, however, did not use the term robot in abstract conceptualizations [2], suggesting variation in how embodiment is represented across contexts.

Within the affective dimension, students' emotional references ranged from curiosity and enjoyment to unease and caution. Positive expressions emphasized AI's usefulness in learning and daily activities, while negative comments centered on fears of dominance or loss of control. This ambivalence is consistent with findings from Dutch and Swedish studies [1,2], where mixed emotional evaluations were also present. Some responses reflected familiar media narratives, such as robots taking over the world, echoing motifs documented in previous European research [1,6]. The coexistence of enthusiasm and concern indicates that students' emotional conceptions of AI draw both on everyday interactions and on fictional portrayals circulating through popular culture.

In the behavioral domain, students described using AI mainly through voice assistants, educational platforms, and games, suggesting exploratory rather than systematic interaction, a pattern comparable to Finnish and Dutch data [2,6]. Some participants mentioned using generative tools for creative purposes, such as producing images or solving tasks, reflecting awareness of recent technological developments. Compared with Dutch students who articulated algorithmic manipulation strategies (e.g., "when you press 'not interested' you will see those videos less") [2,6], Greek students focused more on visible outcomes of AI rather than on underlying algorithms. This contrast may reflect contextual differences in students' exposure to digital recommendation systems or explicit algorithmic language.

Concerning the ethical dimension, students referred to issues such as responsible use, fairness, privacy, and employment. Several mentioned that humans create and can control AI systems, a point also noted in Walan's [1] Swedish dataset. Ethical meanings appeared not only in definitional responses but also in word associations such as responsibility,

https://doi.org/10.3390/computers15010014

rules, respect, and humans, indicating intuitive links between AI and moral or social responsibility. Although brief and largely implicit, these references reveal that ethical reasoning emerges gradually and becomes more articulated during late primary school years, aligning with international findings that children spontaneously mention moral aspects when discussing technology. Comparatively, several cross-national patterns were evident. Anthropomorphic conceptions and notions of AI autonomy recurred across Greek, Swedish, and Turkish contexts [1,6] while Finnish data contained fewer references to learning or adaptation [4]. Educational applications such as AI's ability to answer questions were commonly mentioned in Greece, Turkey, and Sweden [1,7]. Emotional ambivalence—curiosity mixed with apprehension—was also a shared feature across [1,2].

At the same time, certain context-specific features characterized the Greek sample. Students frequently referred to current generative systems (e.g., ChatGPT, Tesla, Neuralink), a tendency less visible in earlier or contemporaneous European datasets. Robotic imagery remained prominent in Greek data, while explicit algorithmic strategies—described by Dutch participants [2]—were not observed. These differences may stem from variation in media exposure, everyday technological environments, or language use surrounding AI.

Finally, the data hint at temporal shifts in children's perceptions of AI. Earlier studies focused largely on voice assistants such as Siri and Alexa [4], whereas more recent datasets—including the present one—show references to generative AI (e.g., ChatGPT), consistent with Walan's [1] observations of similar developments in Sweden. Creative uses (e.g., producing images or videos) also appeared here and in Turkish samples linking AI with art and imagination [7]. Such findings indicate that children's representations of AI evolve alongside technological advances and shifts in the tools embedded in everyday life.

The findings indicate that many students already possess an emerging conceptual understanding of AI, providing a valuable foundation for the design of age-appropriate educational initiatives. Early engagement with AI-related topics can help clarify distinctions between concepts that students often conflate—such as robotics and AI—while illustrating their points of intersection [31]. Schools play a pivotal role in shaping students' AI literacy and in addressing misconceptions before they become entrenched [32]. Students' misunderstandings tend to be superficial rather than fundamental [4], suggesting that accessible explanations of how AI functions, supported by concrete and relatable examples, can enhance comprehension and interest.

Educational programs should therefore emphasize the functional diversity and societal relevance of AI applications, linking them with familiar contexts such as communication, creativity, and learning support. Introducing AI literacy at an early stage can also promote awareness of the human characteristics often attributed to AI, fostering informed reflection rather than uncritical anthropomorphism [33].

The coexistence of positive and negative associations in students' perceptions provides a natural entry point for ethical discussion in the classroom. Encouraging students to examine both the opportunities and limitations of AI aligns with current frameworks of AI literacy [34–36]. The educational objective is not to cultivate uniformly positive views but to develop realistic and critically informed conceptions of AI that enable learners to participate thoughtfully in societal and ethical discussions about its use [37].

6. Conclusions, Limitations, and Future Research

This study provides a qualitative account of how Greek upper primary students conceptualize Artificial Intelligence across cognitive, affective, behavioral, and ethical dimensions prior to any formal instruction. The study contributes new cross-cultural evidence to the international AI literacy literature by examining these four dimensions in a previously underexplored national context. The findings demonstrate that students'

https://doi.org/10.3390/computers15010014

perceptions of AI are multifaceted, blending functional understandings of technology with emotional ambivalence, everyday usage narratives, and spontaneous ethical reflections. This complexity underscores the importance of approaching AI literacy not as a corrective response to misconceptions, but as an educational opportunity to build upon children's already emerging conceptual frameworks.

From an educational perspective, the study highlights the need for human-centered, developmentally appropriate AI curricula that address cognition, emotion, action, and ethics in integrated ways. Students' intuitive references to responsibility, societal impact, and human agency indicate early readiness for critical technological engagement, even at the primary level. AI literacy initiatives should therefore extend beyond technical familiarity to include reflective discussion of the values, limitations, and social consequences of AI technologies. Such holistic approaches can contribute to the cultivation of informed, ethically grounded digital citizens capable of meaningful participation in an AI-shaped society.

As a qualitative exploration, this study aimed to generate context-rich, transferable insights rather than broad generalizations. The sample—comprising students from urban schools in Athens—reflects a specific sociocultural environment that may differ from those in other regions or educational systems. Consequently, future research could adopt comparative or mixed-method designs that include rural, multilingual, or socioeconomically diverse contexts to better understand the variability of students' perceptions. Given the rapid evolution of generative AI tools, it would also be valuable to examine how students' conceptions change over time as they encounter new technologies both inside and outside the classroom.

Methodologically, further work could extend beyond short written responses to incorporate multimodal techniques such as drawings, interviews, and classroom observations. These approaches could provide deeper insight into how children's ideas and emotional responses develop in relation to AI. Additionally, investigating teachers' perceptions and instructional practices would complement student-centered findings, offering a fuller picture of how AI literacy can be meaningfully integrated into primary education.

Taken together, such future studies could inform the design of AI literacy curricula that holistically address the cognitive, emotional, behavioral, and ethical dimensions of students' engagement with AI, fostering balanced, informed, and responsible interaction with intelligent technologies from an early age.

Author Contributions: Conceptualization, K.K., G.C. and P.A.; methodology, G.C.; validation, G.C., K.K. and P.A.; formal analysis, G.C. and K.K.; investigation K.K. and G.C.; resources G.C. and K.K.; writing—original draft preparation, G.C. and K.K.; writing—review and editing, G.C., K.K. and P.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and the research protocol was approved by the Directorate of Primary Education of D' Athens (D.P.E. D' Athens), Ministry of Education, Religious Affairs and Sports (Project identification code: 4830), on 2 May 2025.

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

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

Acknowledgments: The authors would like to thank all participating teachers for their valuable contributions to this study.

https://doi.org/10.3390/computers15010014

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

Abbreviations

The following abbreviations are used in this manuscript:

AI Artificial Intelligence

References

1. Walan, S. Primary School Students' Perceptions of Artificial Intelligence–for Good or Bad. *Int. J. Technol. Des. Educ.* **2025**, *35*, 25–40. [CrossRef]
2. Heeg, D.M.; Avraamidou, L. Young Children's Understanding of AI. *Educ. Inf. Technol.* **2025**, *30*, 10207–10230. [CrossRef]
3. Bewersdorff, A.; Zhai, X.; Roberts, J.; Nerdel, C. Myths, Mis- and Preconceptions of Artificial Intelligence: A Review of the Literature. *Comput. Educ. Artif. Intell.* **2023**, *4*, 100143. [CrossRef]
4. Mertala, P.; Fagerlund, J.; Calderon, O. Finnish 5th and 6th Grade Students' Pre-Instructional Conceptions of Artificial Intelligence (AI) and Their Implications for AI Literacy Education. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100095. [CrossRef]
5. Kim, K.; Kwon, K.; Ottenbreit-Leftwich, A.; Bae, H.; Glazewski, K. Exploring Middle School Students' Common Naive Conceptions of Artificial Intelligence Concepts, and the Evolution of These Ideas. *Educ. Inf. Technol.* **2023**, *28*, 9827–9854. [CrossRef]
6. Kalemkuş, J.; Kalemkuş, F. Primary School Students' Perceptions of Artificial Intelligence: Metaphor and Drawing Analysis. *Eur. J. Educ.* **2025**, *60*, e70007. [CrossRef]
7. Oruc, T.; Korkmaz, O.; Kurt, M. Primary School Students' Views on Artificial Intelligence. *Int. J. Technol. Educ. Sci.* **2024**, *8*, 583–601. [CrossRef]
8. Mitcham, C. *Thinking Through Technology: The Path Between Engineering and Philosophy*; University of Chicago Press: Chicago, IL, USA, 1994.
9. Ankiewicz, P. Alignment of the traditional approach to perceptions and attitudes with Mitcham's philosophical framework of technology. *Int. J. Technol. Des. Educ.* **2018**, *29*, 329–340. [CrossRef]
10. Holmes, W.; Bialik, M.; Fadel, C. *Artificial Intelligence in Education Promises and Implications for Teaching and Learning*; Center for Curriculum Redesign: Boston, MA, USA, 2019.
11. Asghar, M.Z.; Duah, K.A.; Iqbal, J.; Järvenoja, H. Evidence from West Africa on the Interplay of Affective Behavioral Cognitive and Ethical Dimensions of AI Literacy in Ghanaian and Nigerian Universities. *Discov. Comput.* **2025**, *28*, 169. [CrossRef]
12. Blumer, H. What Is Wrong with Social Theory? *Am. Sociol. Rev.* **1954**, *19*, 3–10. [CrossRef]
13. Dey, I. *Qualitative Data Analysis: A User Friendly Guide for Social Scientists*; Routledge: London, UK, 2003.
14. Cadime, I.; Mendes, S.A. Psychological Assessment in School Contexts: Ethical Issues and Practical Guidelines. *Psicol. Reflexão Crít.* **2024**, *37*, 32. [CrossRef]
15. Angelöw, A.; Psouni, E. Participatory Research With Children: From Child-Rights Based Principles to Practical Guidelines for Meaningful and Ethical Participation. *Int. J. Qual. Methods* **2025**, *24*, 16094069251315391. [CrossRef]
16. Fraenkel, J.R.; Wallen, N.E. *How to Design and Evaluate Research in Education*; McGraw-Hill: New York, NY, USA, 1990.
17. Henrich, M.; Formella-Zimmermann, S.; Schneider, S.; Dierkes, P.W. Free Word Association Analysis of Students' Perception of Artificial Intelligence. *Front. Educ.* **2025**, *10*, 1543746. [CrossRef]
18. Braun, V.; Clarke, V. *Thematic Analysis: A Practical Guide*; Sage Publications: London, UK, 2022.
19. Saldaña, J. *The Coding Manual for Qualitative Researchers*; Sage Publications: Thousand Oaks, CA, USA, 2021.
20. Moncada, M. Should We Use NVivo or Excel for Qualitative Data Analysis? *BMS Bull. Sociol. Methodol. Bull. Methodol. Sociol.* **2025**, *165–166*, 186–211. [CrossRef]
21. Corbin, J.; Strauss, A. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*; Sage Publications: Thousand Oaks, CA, USA, 2014.
22. Bingham, A.J. From Data Management to Actionable Findings: A Five-Phase Process of Qualitative Data Analysis. *Int. J. Qual. Methods* **2023**, *22*, 16094069231183620. [CrossRef]
23. Sandelowski, M. Focus on Research Methods Real Qualitative Researchers Do Not Count: The Use of Numbers in Qualitative Research. *Res. Nurs. Heal.* **2001**, *24*, 230–240. [CrossRef]
24. Drisko, J.W. Transferability and Generalization in Qualitative Research. *Res. Soc. Work Pract.* **2025**, *35*, 102–110. [CrossRef]
25. Zaidi, S.B. Situating Sensitizing Concepts in the Constructivist-Critical Grounded Theory Method. *Int. J. Qual. Methods* **2022**, *21*, 16094069211061957. [CrossRef]
26. Lincoln, Y.S.; Guba, E.G. *Naturalistic Inquiry*; Sage: Beverly Hills, CA, USA, 1985.
27. Cohen, L.; Manion, L.; Morrison, K. *Research Methods in Education*; Routledge: London, UK, 2007.

https://doi.org/10.3390/computers15010014

28. O'Connor, C.; Joffe, H. Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *Int. J. Qual. Methods* **2020**, *19*, 1609406919899220. [CrossRef]
29. Lim, W.M. What Is Quantitative Research? An Overview and Guidelines. *Australas. Mark. J.* **2025**, *33*, 325–348. [CrossRef]
30. Novak, A.M.; Jen, E. The Editorial Word: Trustworthiness. *J. Adv. Acad.* **2024**, *35*, 718–728. [CrossRef]
31. Cave, S.; Coughlan, K.; Dihal, K. "Scary Robots" Examining Public Responses to AI. In Proceedings of theAIES 2019-Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, Honolulu, HI, USA, 27–29 January 2019; pp. 331–337. Available online: https://dl.acm.org/doi/proceedings/10.1145/3306618 (accessed on 4 November 2025).
32. Sperling, K.; Stenliden, L.; Mannila, L.; Hallström, J.; Nordlöf, C.; Heintz, F. Perspectives on AI Literacy in Middle School Classrooms: An Integrative Review. *Postdigital Sci. Educ.* **2025**, *7*, 719–749. [CrossRef]
33. Wood, G.; Nuñez Castellar, E.; IJsselsteijn, W. An Exploratory Study Into the Impact of AI Literacy Training on Anthropomorphism and Trust in Conversational AI. *Lect. Notes Comput. Sci.* **2025**, *15820 LNAI*, 301–322. [CrossRef]
34. Kotsidis, K.; Dima, A. Integrating AI Tools and Drama Pedagogy in Digital Classrooms to Foster Critical Thinking and Inclusion in Primary Education. *Adv. Mob. Learn. Educ. Res.* **2025**, *5*, 1524–1535. [CrossRef]
35. Chee, H.; Ahn, S.; Lee, J. A Competency Framework for AI Literacy: Variations by Different Learner Groups and an Implied Learning Pathway. *Br. J. Educ. Technol.* **2025**, *56*, 2146–2182. [CrossRef]
36. Kotsidis, K.; Anastasiades, P. E-Learning Open Seminar on "Human–Centered Artificial Intelligence in Education: From Theory to Practice". *Int. J. Educ. Technol. Learn.* **2025**, *18*, 35–40. [CrossRef]
37. Yue, M.; Jong, M.S.Y.; Dai, Y.; Lau, W.W.F. Students as AI Literate Designers: A Pedagogical Framework for Learning and Teaching AI Literacy in Elementary Education. *J. Res. Technol. Educ.* **2025**, *1*, 1–22. [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.

computers

MDPI

Article

Optimizing Cloudlets for Faster Feedback in LLM-Based Code-Evaluation Systems

Daniel-Florin Dosaru *, Alexandru-Corneliu Olteanu * and Nicolae Țăpuș

Computer Science Department, Faculty of Automatic Control and Computer Science, National University of Science and Technology Politehnica Bucharest, 060042 Bucharest, Romania; nicolae.tapus@upb.ro
* Correspondence: daniel.dosaru@upb.ro (D.-F.D.); alexandru.olteanu@upb.ro (A.-C.O.)

Abstract: This paper addresses the challenge of optimizing cloudlet resource allocation in a code evaluation system. The study models the relationship between system load and response time when users submit code to an online code-evaluation platform, LambdaChecker, which operates a cloudlet-based processing pipeline. The pipeline includes code correctness checks, static analysis, and design-pattern detection using a local Large Language Model (LLM). To optimize the system, we develop a mathematical model and apply it to the LambdaChecker resource management problem. The proposed approach is evaluated using both simulations and real contest data, with a focus on improvements in average response time, resource utilization efficiency, and user satisfaction. The results indicate that adaptive scheduling and workload prediction effectively reduce waiting times without substantially increasing operational costs. Overall, the study suggests that systematic cloudlet optimization can enhance the educational value of automated code evaluation systems by improving responsiveness while preserving sustainable resource usage.

Keywords: cloudlet; automated code evaluation; OnlineJudge; GenAI

1. Introduction

Writing maintainable, high-quality code and receiving prompt, clear feedback significantly enhance learning outcomes [1] in computer science education. As students develop programming skills, rapid, reliable evaluation helps reinforce good practices, correct misconceptions early, and maintain engagement—especially in campus-wide online judge systems. Similarly, in the software industry, timely code review and automated testing play crucial roles in ensuring code quality, reducing defects, and accelerating development cycles, underscoring the importance of feedback-driven workflows in both education and professional practice.

This paper presents optimization strategies for managing cloudlet resources of LambdaChecker, an online code evaluation system developed at the National University of Science and Technology Politehnica Bucharest. The system was designed to enhance both teaching and learning in programming-focused courses such as Data Structures and Algorithms and Object-Oriented Programming. It does this by supporting hands-on activities, laboratory sessions, examinations, and programming contests. A key advantage of LambdaChecker is its extensibility, which enables the integration of custom features tailored to specific exam formats and project requirements. Moreover, the platform facilitates the collection of coding-related metrics, including the detection of pasted code fragments, the recording of job timestamps, and the measurement of the average time required to solve a task.

In its current implementation, LambdaChecker assesses code correctness through input/output testing and evaluates code quality [2] using PMD [3], an open-source static code analysis tool. PMD, which is also integrated into widely used platforms such as SonarQube, provides a set of carefully selected object-oriented programming metrics. These metrics are sufficiently simple to be applied effectively in an educational context while still aligning with industry standards.

In our previous work [4], building on earlier studies [5,6], we extended LambdaChecker to include support for design pattern detection using Generative AI (GenAI). We have used the LLaMA 3.1 model [7] (70.6B parameters, Q4_0 quantization) to detect design patterns in the submitted user's code. This functionality has proven particularly valuable for teaching assistants during the grading process, as it facilitates the assessment of higher-level software design principles beyond code correctness and quality. This integration leverages GenAI models to enhance the reliability of automated feedback while reducing the grading workload. Nonetheless, this advancement is associated with higher resource requirements and increased latency in code evaluation systems without optimization.

During a contest, the workload on our system differs significantly from regular semester activity, exhibiting a sharp increase in submission rates. While day-to-day usage remains relatively irregular and influenced by teaching assistants' instructional styles, content periods generate highly concentrated bursts of activity. The submission rate rises steadily as users progress with their tasks, culminating in peak workloads (e.g., the peak in Figure 1, approximately 54 submissions per minute) during the contest. Such intense demand places substantial pressure on computational and storage resources, making it essential to optimize resource allocation. Without efficient resource management, the platform risks performance degradation or downtime precisely when reliability is most critical. Consequently, adapting the system's infrastructure to match workload patterns dynamically ensures both cost-effectiveness during low-activity periods and robustness during high-stakes contest scenarios.

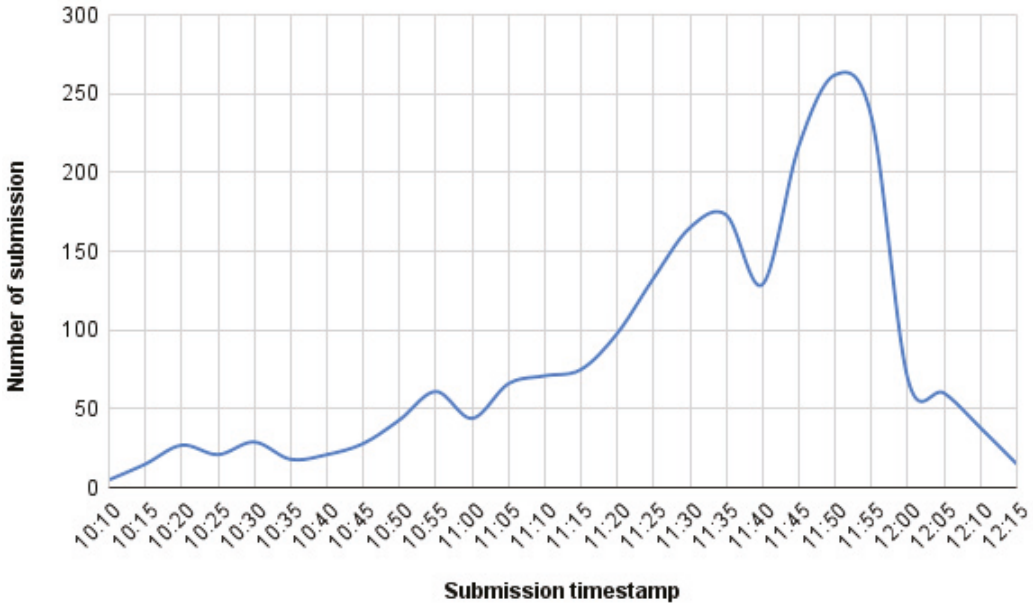


Figure 1. Frequency of submissions in 5 min intervals during the OOP practical examination.

To address the resource inefficiencies observed in the previous version of LambdaChecker with LLM-based design pattern detection [4], we employed optimization strategies to manage computational resources within the cloudlet. Our main contributions are:

- Mathematical modeling of resource allocation: We develop a model for dynamically allocating computational resources to ensure stable performance during contests and regular lab activities.
- Simulation-based validation: We evaluate the model under high submission rates using real contest data, demonstrating reduced latency, improved throughput, and enhanced stability while remaining cost-effective.
- Integration with LLM-assisted feedback: We demonstrate that LLM-powered design pattern detection can coexist with optimized resource management without compromising responsiveness.

The paper proceeds as follows: Section 2 reviews the related work in the field, while Section 3 describes the cluster, the LambdaChecker system, and the mathematical model and optimization techniques. Section 4 applies the mathematical model in a case study—LambdaChecker during a contest. Section 5 contains the discussion and user experience implications, and Section 6 concludes the paper.

2. Related Work

Automated code evaluation systems, commonly known as online judges (OJs), play a central role in both competitive programming and computer science education. Early research demonstrated the feasibility of cloud-based architectures for scalable OJ platforms, enabling the simultaneous evaluation of many user submissions [8]. Subsequent work leveraged container-based sandboxing to improve isolation and resource efficiency, allowing platforms to handle the bursty workloads typical of exams and practical assignments [9]. Surveys of automated feedback systems in education further emphasize the importance of combining correctness checking, static analysis, and pedagogically meaningful feedback to enhance learning outcomes [10,11]. These studies establish the foundational need for efficient, responsive evaluation pipelines, particularly in time-constrained assessment scenarios.

Cloudlet and edge computing approaches have been proposed to improve responsiveness and resource utilization in latency-sensitive applications. Task offloading and collaborative edge–cloud scheduling strategies demonstrate that local compute layers can significantly reduce response times while balancing load across distributed infrastructure [12]. Surveys of edge networks further highlight how the convergence of computation, caching, and communication improves performance under highly variable workloads [13]. To handle fluctuating demand, both classical and modern auto-scaling strategies have been studied, ranging from hybrid cloud auto-scaling techniques aimed at cost-performance tradeoffs [14] to reinforcement-learning-based adaptive scheduling [15]. These insights inform the design of short-, medium-, and long-term scheduling strategies for cloudlet-based code evaluation systems, such as the one presented in this study.

Analytical performance modeling continues to provide guidance for managing high-volume workloads in interactive and multi-tier systems. Recent efforts have extended classical queueing and simulation-based modeling to modern microservice or cloud-native platforms. For example, PerfSim: A Performance Simulator for Cloud Native Microservice Chains offers a discrete-event simulation tool that, given traces of microservice behavior, can simulate entire service chains and predict key metrics such as average response times under different resource allocation policies [16]. Similarly, a performance modeling framework for microservices-based cloud infrastructures proposes a comprehensive model tailored to microservice-based cloud systems that supports capacity planning and resource provisioning decisions under realistic deployment conditions [17]. Moreover, acknowledging non-stationarity in arrival and service patterns, performance modeling of cloud

systems by an infinite-server queue operating in a rarely changing random environment uses queueing theory with a random-environment extension to reflect dynamic workload fluctuations [18]. These modern approaches support the use of queueing-based simulations, microservice-aware modeling, and stochastic workload modeling as essential tools for optimizing cloudlet behavior in automated code evaluation pipelines.

Recent advances in large language models (LLMs) have enabled novel forms of automated code analysis and feedback. Generative AI models have been used to assess code quality, detect design patterns, and provide higher-level feedback beyond simple correctness checks [5]. Evaluations of code-oriented LLMs demonstrate their ability to classify, analyze, and predict defects across various programming languages and tasks [6]. Integrating a local LLM into a cloudlet-based evaluation pipeline allows for richer analysis—such as design-pattern detection—without sacrificing latency or data privacy, complementing traditional static analysis tools like PMD.

Quick, targeted feedback can significantly help students learn to code. Adaptive feedback that addresses specific errors allows beginners to correct mistakes more efficiently and supports iterative problem-solving [19]. Research shows that automated feedback improves performance on assignments, particularly for more complex problems, though its benefits may be less pronounced when feedback is no longer provided. In line with this, the implementation of Quick Check questions and increased exam proctoring provides more immediate feedback and structured engagement opportunities. These shorter feedback cycles appear to enhance learning by promoting consistent interaction with course materials, improving performance on application-based questions, and fostering deeper cognitive processing [20]. This highlights that optimizing response times in online code evaluation systems is not just a matter of technical efficiency—providing faster and more consistent feedback directly supports learning. It creates a more effective and engaging user experience.

Govea et al. [21] explore how artificial intelligence and cloud computing can be integrated to improve the scalability and performance of educational platforms. Their work highlights strategies for efficient management of computational resources under high user loads, providing insights directly relevant to optimizing cloud utilization in student code evaluation systems. However, their approach is more general, targeting educational platforms broadly rather than focusing specifically on code evaluation. As a result, it lacks tailored mechanisms for handling programming assignments, such as automated testing, code similarity detection, or fine-grained feedback, which are central to dedicated student code evaluation systems.

Kim et al. [22] present Watcher, a cloud-based platform for monitoring and evaluating student programming activities. The system emphasizes isolation and convenience, providing each student with an independent virtual machine to prevent interference and ensure secure execution of assignments. By capturing detailed coding activity logs and enabling instructors to analyze potential cheating or errors, Watcher demonstrates how cloud-based infrastructure can support fair, scalable, and secure student code evaluation. Nevertheless, relying on dedicated virtual machines for each student can lead to higher resource usage and operational costs, potentially constraining scalability in large classes or institutions with limited cloud budgets. Furthermore, this system does not account for the substantial resource requirements of LLM-based evaluation.

Queueing-theory models developed for fog computing provide valuable insights for managing low-latency, distributed resources similar to cloudlets. For example, a recent study presents a queuing model for fog nodes that captures task arrival rates, service times, and resource contention to predict system performance under dynamic workloads [23]. Although their focus is on IoT and general edge applications, the principles directly apply to

cloudlet-based code-evaluation systems, where jobs from multiple users must be processed efficiently while minimizing feedback latency. Incorporating such analytical models enables us to reason about throughput, average response times, and system utilization, which form the basis for our optimization of job scheduling and resource allocation in a sequential, multi-user environment.

While each of these studies provides valuable insights, none fully addresses the combined challenges of sequential job evaluation, fine-grained feedback, and cost-effective cloud utilization in the context of real-time LLM-based code evaluation systems. Our work builds on these foundations by developing a tailored mathematical model for load and response time, specifically designed for user code evaluation, estimated workload, and by analyzing optimization strategies that balance performance, cost, and user performance outcomes. This positions our approach as a bridge between theoretical cloud optimization, practical monitoring systems, and pedagogically meaningful evaluation.

3. Materials and Methods

In the initial iterations of the system, we observed the critical role played by platform performance and students' sensitivity to the overhead introduced by the tooling. When integrating the LLM and PMD components, our primary concern was to avoid increasing feedback time. This constraint directly influenced the choice and applicability of the proposed model to the LambdaChecker data, as it reflects real-world performance requirements observed empirically during early deployments.

3.1. Cluster Setup, LLM-Based Code Evaluation Tool

Owing to limited computational resources, we initially assigned all submission-processing tasks for the LambdaChecker scheduler to a dedicated compute node. This node is equipped with Intel® Xeon® E5-2640 v3 multicore processor two NVIDIA H100 GPUs with 80 GB of HBM3 memory each, providing substantial memory capacity and computational power. This setup is ideal for large-scale models; for instance, a pair of these GPUs can efficiently run a 70B-parameter model like LLaMA 3.1 while maintaining high performance and optimized memory usage.

The machine supports the entire evaluation pipeline, handling CPU-intensive tasks such as PMD static analysis and correctness testing, as well as GPU-accelerated design-pattern detection workloads. Each step increases the load on the evaluation system, with LLM-based design pattern detection being the most resource-intensive, consuming up to 70% of the total average evaluation time.

Figure 2 presents the architecture of the LambdaChecker queue-driven code evaluation system deployed within the cloudlet environment. Users submit source code via a web-based frontend that communicates with a backend API responsible for managing submission requests and responses. Each submission is encapsulated as a job and placed into a centralized job queue within the cloudlet. A dynamically scaled pool of worker virtual machines retrieves jobs from the queue. It executes an evaluation pipeline that begins with test case execution, then performs object-oriented code quality analysis with PMD, and concludes with software design pattern detection by the LLaMA 3.1 large language model. Once the entire evaluation pipeline is complete, the output is retrieved by the backend API and presented to the user through the frontend interface. System-level monitoring continuously observes job-queue metrics, which are fed into an auto-scaling subsystem that provisions or deprovisions worker VMs in response to workload fluctuations, ensuring scalable throughput and low-latency processing under varying demand conditions.

Evaluation tasks include precedence constraints within the worker pipeline. Each step depends on the successful completion of the previous one: test cases run first, followed by

PMD analysis, and then LLaMA pattern detection. If test cases fail or take longer than the timeout time, subsequent steps are skipped.

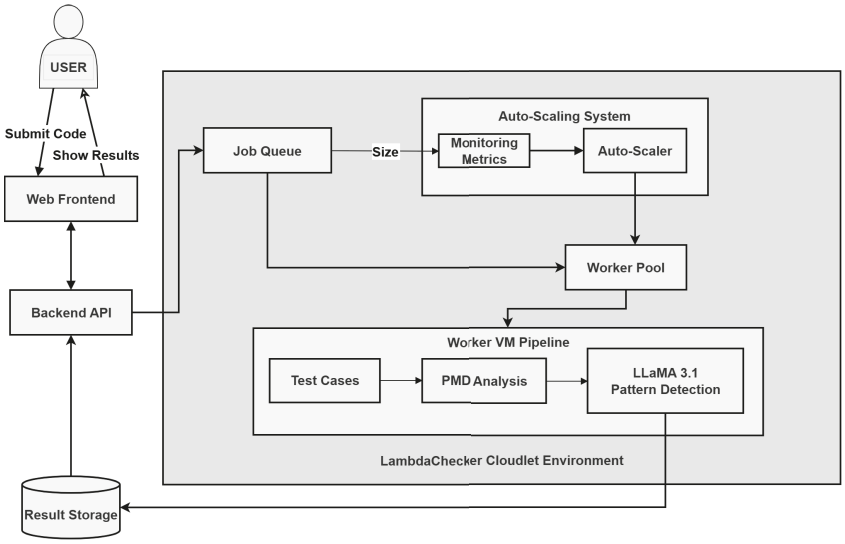


Figure 2. LambdaChecker queueing system in the cloudlet environment.

A cloudlet-based LLM offers three main advantages over a traditional cloud deployment: (1) lower latency, since computation occurs closer to the user and enables faster, more interactive code analysis; (2) improved data privacy, as sensitive source code remains within a local or near-edge environment rather than being transmitted to distant cloud servers; and (3) reduced operational costs, because frequent analysis tasks avoid cloud usage fees and bandwidth charges while still benefiting from strong computational resources.

User-submitted code is incorporated into a pre-defined, improved prompt tailored to the specific coding problem to facilitate the detection of particular design patterns. The large language model then performs inference on a dedicated GPU queue supported by an NVIDIA H100 accelerator (NVIDIA Corporation, Santa Clara, CA, USA).

The prompt was refined to be clearer, more structured, and easier for a generative AI model to follow. It explicitly defines the JSON output, clarifies how to report missing patterns, and instructs the model to focus on real structural cues rather than class names. The rule mapping publisher–subscriber to the Observer pattern ensures consistent terminology. Overall, the prompt improves reliability and reduces ambiguity in pattern detection.

The prompt is concise, focuses only on essential instructions, and specifies a strict JSON output format, allowing the cloudlet-hosted LLM to process the code more efficiently without unnecessary reasoning or verbose explanations:

"You analyze source code to identify design patterns. Output ONLY this JSON:

```
  {
  "design_patterns": [
    {
      "pattern": "<pattern_name>",
      "confidence": <0-100>,
      "adherence": <0-100>
    }
  ]
}
```

If no patterns are found, return {"design_patterns": []}."

Using this prompt, we obtained the following precision results. As previously reported in [4], the LLaMA model exhibited varying effectiveness in detecting design patterns on a separate dataset of student homework submissions. In that study, the model achieved high precision for Singleton (1.00) and Visitor (0.938), moderate precision for Factory (0.903) and Strategy (0.722), and lower precision for Observer (0.606). Recall values ranged from 0.536 for Visitor to 0.952 for Observer, highlighting differences in the model's performance across patterns. We include the model overall results in Table 1.

Table 1. Overall results for the LLaMA model on design pattern detection.

Metric	**LLaMA**
Precision	0.779
Recall	0.769
F1-score	0.774

The system dynamically scales across multiple computing nodes in response to the number of active submissions. It continuously monitors the average service rate, which depends on the current submissions workload, and adjusts resources in a feedback-driven loop, scaling nodes up or down as required to ensure efficient processing. The scaling mechanism is based on a mathematical model described in the following subsection.

3.2. Mathematical Model

In our analysis, we consider only submissions that successfully pass all test cases. Submissions that fail or exceed the allowed time at any stage are eliminated, ensuring that only items which progress through all three stages of the pipeline (Test Cases $\rightarrow$ PMD Analysis $\rightarrow$ LLaMA Pattern Detection) are included. This allows us to measure the complete processing time W for each item consistently.

Additionally, there is no serial dependency beyond per-user ordering. Tasks from different users are independent and can run concurrently, so the system-level analysis assumes parallel execution across users, while each user's pipeline remains sequential.

We consider a generic queueing system and introduce the following notation:

- $L(t)$: average queue length at time t;
- $\lambda(t)$: arrival rate at time t (jobs per unit time);
- $\lambda_{\text{eff}}(t)$: effective arrival rate at time t (jobs per unit time);
- μ: service rate per machine (jobs processed per unit time);
- m: number of parallel processing machines (queues).

Even if the exact inter-arrival and service-time distributions are unknown, empirical traces in real systems typically exhibit diurnal variability, burstiness near deadlines, and transient queue-drain phases. A useful analytical approximation is to treat the system as an *M/M/m* queue with effective rates

$$\mu_{\text{eff}} = m\mu, \qquad \lambda_{\text{eff}}(t) \leq \lambda(t), \tag{1}$$

leading to an approximate expected queue length

$$L(t) \approx \frac{\lambda_{\text{eff}}(t)}{\mu_{\text{eff}} - \lambda_{\text{eff}}(t)}. \tag{2}$$

This approximation is valid whenever the stability condition

$$\lambda_{\text{eff}}(t) < \mu_{\text{eff}} \tag{3}$$

is satisfied; otherwise, the system becomes overloaded, and the queue diverges.

Although submission traces (Figure 1) exhibit burstiness and deviations from Poisson arrivals, we use the *M/M/m* model as a first-order approximation to capture the average system load. While the Poisson/exponential assumptions do not strictly hold, this approximation is valuable because it provides analytical insight into expected queue lengths and utilization. To partially account for burstiness, we introduce an *effective arrival rate* $\lambda_{\text{eff}}(t) \leq \lambda(t)$, which smooths short-term bursts and avoids overestimating congestion. The approximate expected queue length is then the one presented in the equations above.

For periods when $\lambda(t) > \mu_{\text{eff}}$ and the system is overloaded, no stationary distribution exists. In this case, we use a fluid approximation that captures average queue growth:

$$\frac{dL}{dt} = \lambda - \mu_{\text{eff}}. \tag{4}$$

If arrivals are indexed deterministically so that the n-th submission occurs at

$$t_n \approx \frac{n}{\lambda}, \tag{5}$$

then the approximate waiting time experienced by the n-th job is as follows:

$$W_n \approx \frac{L(t_n)}{\mu} = \frac{\lambda - \mu}{\lambda\mu}\, n. \tag{6}$$

Limitations: This model may underestimate peak queue lengths and does not accurately capture tail latencies. It is therefore intended primarily for average-load analysis rather than for fine-grained delay predictions.

4. Results

4.1. Case Study: LambdaChecker in a Programming Contest Increasing the Service Rate

We begin by analyzing a highly loaded contest scenario through a hypothetical *M/M/1* queueing model.

We take these values from our dataset. As shown in Figure 1, we observe an average peak of approximately 270 submissions within 5-min intervals during the OOP practical examination. This gives an arrival rate of

$$\lambda \approx \frac{270\ \text{jobs}}{5\ \text{min}} = 54\ \text{jobs/min}.$$

After analyzing a historical contest dataset, the total evaluation time for submitted code ranges from 6.5 s to 17 s, with an average duration of 12.02 s. The code correctness evaluation is strictly limited to 2 s; executions that exceed this threshold result in a timeout exception and are excluded from further analysis. The PMD code quality analysis has a relatively fixed execution time of approximately 1.5 s. The remaining portion of the evaluation time is dominated by LLM inference, which varies with the input size (i.e., the number of tokens) and takes approximately 8.5 s on average. Given the average total evaluation time, this corresponds to a service rate of

$$\mu = \frac{1}{12.02\ \text{s}} \times 60\ \frac{\text{s}}{\text{min}} \approx 4.99\ \text{jobs/min}.$$

The traffic intensity is

$$\rho = \frac{\lambda}{\mu} = \frac{54}{4.99} = 10.82,$$

indicating severe overload and unbounded queue growth.

If we consider a single queue system, we can compute the final submission with $n = 2119$ its waiting time:

$$\begin{aligned}
\lambda - \mu &= 49.01, \\
\lambda\mu &= 269.46, \\
\frac{\lambda - \mu}{\lambda\mu} &= \frac{49.01}{269.46} \approx 0.1818, \\
W_{2119} &\approx 0.1818 \cdot 2119 \approx 385 \text{ min} .
\end{aligned}$$

Thus, the 2119th job waits roughly 6 h and 25 min before service begins.

4.2. Determining the Required Number of Machines

To prevent unbounded queue growth during peak contest activity, we model LambdaChecker as an M/M/c system with c identical processing nodes.

Peak Load Parameters (from our dataset)

$$\lambda = 54 \text{ jobs/min}, \qquad \mu = 4.99 \text{ jobs/min per machine}.$$

Total Service Capacity

$$\mu_{\text{total}} = c\mu.$$

Stability Condition

$$\rho = \frac{\lambda}{\mu_{\text{total}}} \leq 1 \quad \Longrightarrow \quad \frac{\lambda}{c\mu} \leq 1.$$

Minimum Machines Required

$$c \geq \frac{\lambda}{\mu} = \frac{54}{4.99} \approx 10.8 \approx 11$$

Thus, at least 11 machines are needed to sustain the peak arrival rate without unbounded queueing. However, $c = 11$ puts the system close to critical load ($\rho = 1$), meaning small fluctuations can still cause delays. In practice, provisioning at least 12 machines yields significantly more stable performance.

By constraining the number of jobs permitted per user and discarding any excess submissions, the system's effective load is reduced, thereby decreasing the required number of nodes. This represents just one specific instance of a broader class of resource-management strategies commonly employed to limit system load. Such methods, while effective at stabilizing resource usage, may result in discarded submissions; thus, an appropriate balance between system robustness and user experience must be carefully maintained.

5. Discussion

We evaluated the impact of our optimization strategies, based on our mathematical model, on system performance as the number of processing nodes increased. In addition, we discuss applying queue policies to limit the number of submissions per user. To validate and explore the queuing behavior empirically, we implemented a discrete-event simulation in Python [24].

In our simulation, arrivals enter a single global queue managed in first-come, first-served (FCFS) order by arrival time. When a processing node becomes free, the job at

the head of the global queue is assigned to that node for immediate processing. Thus, the system operates as a single pooled resource. This discipline corresponds exactly to the standard $M/M/m$ queuing model used in Section 3.2, which assumes a pooled set of identical servers with a shared waiting line. This ensures that no VMs remain idle while jobs wait in the queue. The results are visualized in Figures 3 and 4—in plots of wait times per submission, providing a clear picture of how increasing the number of queues improves performance.

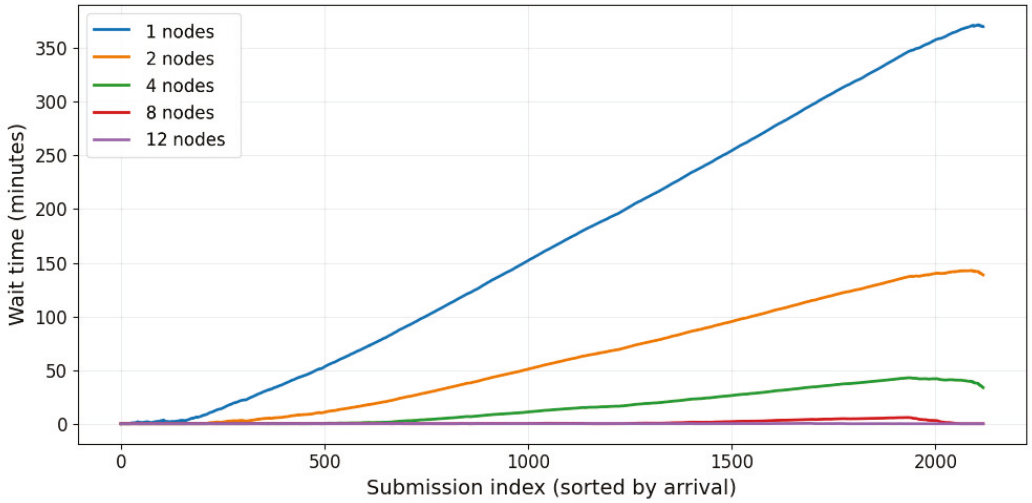


Figure 3. Per-submission wait times across varying numbers of nodes, illustrating the impact of node availability on processing delays. We prove that for our case study, using 12 nodes makes the system stable.

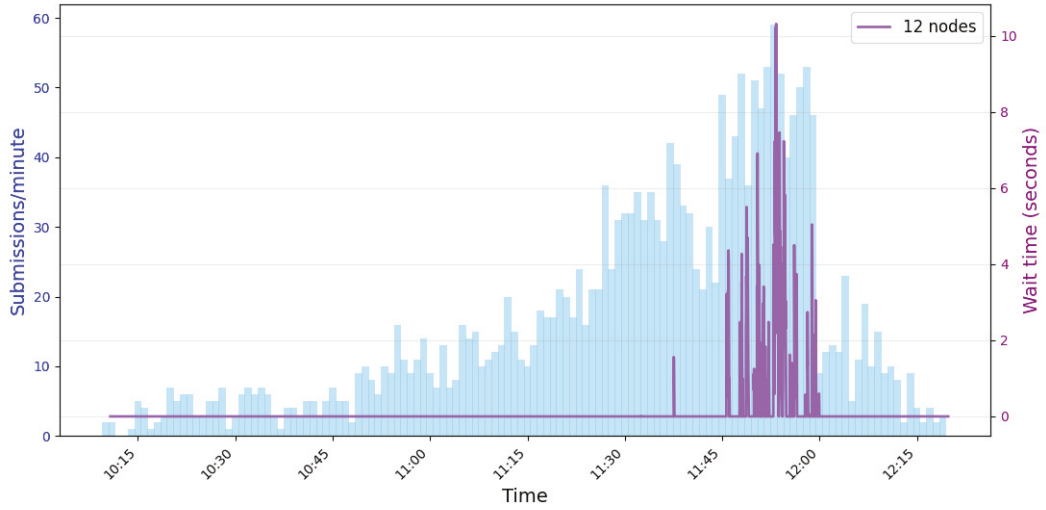


Figure 4. Submission frequency and wait times with a fixed number of 12 processing nodes.

5.1. Effect of Scaling Machines

Figure 3 illustrates the expected waiting time $W(t)$ for different numbers of machines (processing nodes). Consistent with *M/M/m* queueing theory, increasing the number of nodes reduces the waiting time, particularly during peak submission periods. For the original single-server setup, the traffic intensity $\rho \gg 1$, resulting in extremely long queues—our calculations show that the last submissions could experience waiting times exceeding 6 h. Doubling or appropriately scaling the number of machines significantly reduces $W(t)$ and stabilizes the queue. While effective, this approach requires

additional infrastructure, highlighting the trade-off between system performance and operational costs.

In Figure 4, we include a simulation with 12 processing nodes to validate the accuracy of our previous computations. Among the 2119 submissions, the maximum queueing delay observed is approximately ~10.3 s. With 12 processing nodes, the queueing delay remains below 11 s throughout the contest. Given a maximum job evaluation time of approximately 15 s, the resulting end-to-end user-perceived latency, defined as the sum of queueing delay and evaluation time, remains below 25 s. In addition, we overlay the same submission frequency shown in Figure 1 over the contest duration to visualize the submission workload on LambdaChecker for a contest with 200 participants.

5.2. Effect of Queue Policies

By limiting the number of jobs per user and discarding excess submissions, the system reduces load and lowers the number of required nodes. This is an example of a common resource-management strategy that stabilizes usage but may discard submissions, necessitating a careful balance between system robustness and user experience.

5.3. Contest Insights from Code Evaluation System

Several insights arise from our analysis. First, limiting the effective arrival rate is especially impactful when the system is heavily loaded. Due to the non-linear relationship between waiting time and traffic intensity, small reductions in λ_{eff} can lead to disproportionately large decreases in $W(t)$. Second, adding new processing nodes improves performance but increases operational costs, whereas queue policies provide significant benefits at minimal expense. Finally, controlling submissions not only reduces average waiting times but also stabilizes the queue, leading to more predictable processing times for users.

Beyond system metrics, these strategies have a direct impact on user experience and performance. When feedback was delayed or unavailable during the content, participants experienced uncertainty and frustration, often resulting in lower engagement and lower scores. In the initial run, some students waited up to five minutes to receive feedback. In contrast, when feedback was returned more quickly (e.g., in under 25 s), users could identify and correct errors promptly, boosting both learning and confidence. To ensure comparability, participants were offered problems of similar difficulty across conditions. Our data show that with faster feedback, the average score increased from 53 to 69 out of 100. This demonstrates that system-level optimizations not only improve operational efficiency but also meaningfully enhance users' outcomes and engagement.

While several alternative queueing strategies exist in the literature, such as Shortest Job First (SJF) [25] and Multilevel Feedback Queues (MLFQ) [26], these approaches are typically motivated by environments with high variability in job runtimes, preemption support, or runtime estimates. In contrast, student submissions in our system have short, tightly bounded evaluation times (typically 6–18 s) and are executed non-preemptively. Under these conditions, a First-Come-First-Served (FCFS) policy provides a simple, fair, and effective scheduling strategy without the overhead or assumptions required by more complex schedulers. In addition, we incorporate LLM-based design pattern detection to accelerate evaluation and improve accuracy.

5.4. Limitations and Future Work

The *M/M/m* approximation assumes exponential inter-arrival and service times, which may not fully capture the real-world behavior of submissions. Variability in machine performance and user submission patterns may affect the observed service rate μ, and queue-limiting policies could influence user behavior in ways not captured by this static

analysis. Despite these limitations, our combined approach provides a practical and data-driven strategy for managing waiting times effectively, balancing system performance, stability, and operational cost. Our mathematical model may apply to other institutions or online judge platforms with different user populations, submission behaviors, or infrastructure. Future work could explore more realistic arrival distributions and system configurations to capture these practical considerations better.

While our results demonstrate significant performance improvements over our previous system, we did not perform a direct experimental comparison against queueing disciplines such as SJF or MLFQ. We acknowledge this as a limitation of the current study and identify the empirical evaluation of these alternative scheduling policies as a subject for future work.

6. Conclusions

Our analysis shows that a combination of server scaling and queue-control policies can effectively manage system performance during peak submission periods. Analytical calculations indicate that a single-server setup becomes severely overloaded under peak demand, leading to waiting times of several hours. Adding new processing nodes reduces waiting times, although it incurs additional infrastructure costs.

Queue-management policies that limit the number of active submissions per user substantially reduce the effective arrival rate, decreasing both the average waiting time and the queue variability. This approach delivers significant performance improvements at minimal cost and helps stabilize the system, preventing extreme delays even during high submission rates.

The combination of server scaling and submission control yields the best results, maintaining manageable queues and near-optimal waiting times throughout peak periods. Timely feedback directly improves participants' experience and performance. Delays create uncertainty and frustration, lowering engagement, while feedback within 25 s lets users quickly correct errors, boosting understanding, confidence, and contest results. We found that providing faster feedback boosted average scores by approximately 30% on tasks of similar difficulty, highlighting how reducing wait times can improve performance outcomes. These findings demonstrate that system-level optimizations not only improve operational efficiency but also meaningfully enhance users' outcomes and engagement. This emphasizes the importance of balancing technical performance with participants' satisfaction.

Author Contributions: Conceptualization, D.-F.D. and A.-C.O.; methodology, N.Ț.; software, D.-F.D.; validation, A.-C.O. and N.Ț.; formal analysis, D.-F.D.; investigation, D.-F.D. and A.-C.O.; resources, N.Ț.; data curation, D.-F.D.; writing—original draft preparation, D.-F.D.; writing—review and editing, A.-C.O. and N.Ț.; visualization, D.-F.D.; supervision, N.Ț.; project administration, N.Ț.; funding acquisition, A.-C.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Program for Research of the National Association of Technical Universities under Grant GNAC ARUT 2023.

Institutional Review Board Statement: This study utilized fully anonymized data with no collection of personally identifiable information. Based on these characteristics, the study qualified for exemption from Institutional Review Board (IRB) review.

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

Data Availability Statement: Paper dataset and scripts can be found at https://tinyurl.com/2z4tf4yc (accessed on 10 December 2025).

Acknowledgments: The authors would like to thank Emil Racec for coordinating the development of LambdaChecker.

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

References

1. Kuklick, L. When Computers Give Feedback: The Role of Computer-Based Feedback and Its Effects on Motivation and Emotions. *IPN News*, 24 June 2024. Available online: https://www.leibniz-ipn.de/en/the-ipn/current/news/when-computers-give-feedback-the-role-of-computer-based-feedback-and-its-effects-on-motivation-and-emotions (accessed on 10 December 2025).
2. Dosaru, D.-F.; Simion, D.-M.; Ignat, A.-H.; Negreanu, L.-C.; Olteanu, A.-C. A Code Analysis Tool to Help Students in the Age of Generative AI. In Proceedings of the European Conference on Technology Enhanced Learning, Krems, Austria, 16–20 September 2024; pp. 222–228.
3. PMD. An Extensible Cross-Language Static Code Analyzer. Available online: https://pmd.github.io/ (accessed on 10 December 2025).
4. Dosaru, D.-F.; Simion, D.-M.; Ignat, A.-H.; Negreanu, L.-C.; Olteanu, A.-C. Using GenAI to Assess Design Patterns in Student Written Code. *IEEE Trans. Learn. Technol.* **2025**, *18*, 869–876. [CrossRef]
5. Bavota, G.; Linares-Vásquez, M.; Poshyvanyk, D. Generative AI for Code Quality and Design Assessment: Opportunities and Challenges. *IEEE Softw.* **2022**, *39*, 17–24.
6. Chen, M.; Tworek, J.; Jun, H.; Yuan, Q.; de Oliveira Pinto,H.P.; Kaplan, J.; Edwards, H.; Burda, Y.; Joseph, N.; Brockman, G.; et al. Evaluating Large Language Models Trained on Code. *arXiv* **2021**, arXiv:2107.03374. [CrossRef]
7. Meta AI. Introducing Meta Llama 3.1: Our Most Capable Models to Date. 23 July 2024. Available online: https://ai.meta.com/blog/meta-llama-3-1/ (accessed on 10 December 2025).
8. Lu, J.; Chen, Z.; Zhang, L.; Qian, Z. Design and Implementation of an Online Judge System Based on Cloud Computing. In Proceedings of the IEEE 2nd International Conference on Cloud Computing and Big Data Analysis, Chengdu, China, 28–30 April 2017; pp. 36–40.
9. Singh, A.; Sharma, T. A Scalable Online Judge System Architecture Using Container-Based Sandboxing. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 245–252.
10. Keuning, H.; Jeuring, J.; Heeren, B. A Systematic Literature Review of Automated Feedback Generation for Programming Exercises. *ACM Trans. Comput. Educ. (TOCE)* **2018**, *19*, 1–43. [CrossRef]
11. Frolov, A.; Buliaiev, M.; Sandu, R. Automated Assessment of Programming Assignments: A Survey. *Inform. Educ.* **2021**, *20*, 551–580.
12. Li, X.; Tang, W.; Yuan, Y.; Li, K. Dynamic Task Offloading and Resource Scheduling for Edge-Cloud Collaboration. *Future Gener. Comput. Syst.* **2019**, *95*, 522–533.
13. Wang, S.; Zhang, X.; Zhang, Y.; Wang, L.; Yang, J.; Wang, W. A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications. *IEEE Access* **2020**, *8*, 197689–197709. [CrossRef]
14. Spyridon Chouliaras and Stelios Sotiriadis. .An adaptive auto-scaling framework for cloud resource provisioning. *Future Gener. Comput. Syst.* **2023**, *148*, 173–183. [CrossRef]
15. Zhang, Q.; He, Q.; Chen, W.; Chen, S.; Xiang, Y. Adaptive Autoscaling for Cloud Applications via Reinforcement Learning. *IEEE Trans. Cloud Comput.* **2021**, *9*, 1162–1176.
16. Khan, M.G.; Taheri, J.; Al-Dulaimy, A.; Kassler, A. PerfSim: A Performance Simulator for Cloud Native Microservice Chains. *IEEE Trans. Cloud Comput.* **2021**, *18*, 869–876.
17. da Silva Pinheiro, T.F.; Pereira, P.; Silva, B.; Maciel, P. R. A performance modeling framework for microservices-based cloud infrastructures. *J. Supercomput.* **2022**, *79*, 7762–7803. [CrossRef]
18. Moiseeva, S.; Polin, E.; Moiseev, A.; Sztrik, J. Performance Modeling of Cloud Systems by an Infinite-Server Queue Operating in Rarely Changing Random Environment. *Future Internet* **2025**, *17*, 462. [CrossRef]
19. Ahmed, U. Z.; Srivastava, N.; Sindhgatta, R.; Karkare, A. Characterizing the pedagogical benefits of adaptive feedback for compilation errors by novice programmers. In Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET), Seoul, Republic of Korea, 27 June–19 July 2020; pp. 139–150.
20. Law, Y.K.; Tobin, R.W.; Wilson, N.R.; Brandon, L.A. Improving student success by incorporating instant-feedback questions and increased proctoring in online science and mathematics courses. *J. Teach. Learn. Technol.* **2020**, *9*, 1. [CrossRef]
21. Govea, J.; Ocampo Edye, E.; Revelo-Tapia, S.; Villegas-Ch, W. Optimization and scalability of educational platforms: Integration of artificial intelligence and cloud computing. *Computers* **2023**, *12*, 223. [CrossRef]
22. Kim, Y.; Lee, K.; Park, H. Watcher: Cloud-based coding activity tracker for fair evaluation of programming assignments. *Sensors* **2022**, *22*, 7284. [CrossRef] [PubMed]

23. Mas, L.; Vilaplana, J.; Mateo, J.; Solsona, F. A Queuing Theory Model for Fog Computing. *J. Supercomput.* **2022**, *78*, 11138–11155. [CrossRef]
24. Dosaru, D. *Optimized Code Evaluation Simulation*; GitHub Repository, Main Branch. Available online: https://github.com/dosarudaniel/optimized_code_evaluation_simulation/tree/main (accessed on 10 December 2025). Software version used: Python script as available in the `main` branch at the time of access.
25. Alworafi, M.A.; Dhari, A.; Al-Hashmi, A.A.; Darem, A.B.; Suresha. An improved SJF scheduling algorithm in cloud computing environment. In Proceedings of the 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), Mysuru, India, 9–10 December 2016; pp. 208–212.
26. Khan, A.; Khan, M.A.; Kim, S.W. Exploring multilevel feedback queue combinations and dynamic time quantum adjustments. *J. Inf. Sci. Eng.* **2020**, *36*, 1045–1063.

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.

computers

MDPI

Article

Digital Transformation in Design Education: Exploring the Challenges and Opportunities in Jordanian Higher Education

Islam A. Alshafei [1,*], Samah Mohammed AlDweik [2], Mahmoud ali Hassouneh [3], Hanan AbuKarki [4], Abdellatif A. Jarrar [5] and Qusai S. Mansour [6]

[1] Department of Architectural Engineering, Faculty of Engineering, Jerash University, Jerash 26150, Jordan
[2] Faculty of Engineering and Design, Department of Interior Design, Middle East University, Amman 11831, Jordan; sdweik@meu.edu.jo
[3] Faculty of Architecture and Design, Al-Ahliyya Amman University, Amman 19328, Jordan; m.hassouneh@ammanu.edu.jo
[4] Department of Interior Design, Faculty of Architecture & Design, Philadelphia University, P.O. Box 1, Amman 19392, Jordan; hkarki@philadelphia.edu.jo
[5] Applied Art Faculty, National University College of Technology, Amman 22110, Jordan
[6] Faculty of Engineering and Design, Department of Graphic Design, Middle East University, Amman 11831, Jordan; qmansour@meu.edu.jo
* Correspondence: i.alshafei@jpu.edu.jo

Abstract: In recent years, design education has experienced major changes as the number of digital tools and technologies has rapidly developed. Many design programs encounter difficulties in integrating these innovations, despite their potential benefits. In this research, the adoption of digital tools in the teaching of design in Jordanian universities is explored, focusing on the views of educators in relation to their use, the challenges associated with it, and the resultant effects on the pedagogical process. Faculty members working in various departments of design were surveyed gauging the frequency of usage of tools, their knowledge of new technologies, their perceptions of the potential results of an educational process, and the barriers that were met during the integration process. To guide the analysis, three theoretical frameworks were applied: the SAMR model of technology integration, Bloom's Digital Taxonomy, and the Technology Acceptance Model (TAM). The findings reveal that while traditional tools like AutoCAD and Revit are predominantly used at the Substitution and Augmentation stages, emerging technologies such as VR/AR and AI show potential for higher-order integration. However, barriers related to ease of use and perceived usefulness limit their broader adoption. The study contributes to the understanding of digital transformation in design education and provides insights into the pedagogical implications for future curriculum development. The research highlights the need to invest more in the professional development of educators and to work more closely with the technological industry. The proposed implications of these insights concern the restructuring of design education to reflect the needs of the digital age and provide approaches to overcoming obstacles to the successful adoption of technology in teaching environments.

Keywords: digital tools; design education; technology integration; Jordanian universities

1. Introduction

Due to the rapid pace of change in technology, the studies of design are experiencing significant change in the digital era [1]. The introduction of digital tools and platforms into design education has radically changed the practices and methods of design learning

and practice [2,3]. Evidence shows that digital transformation has significant implications for higher education, and specifically in design fields, it introduces opportunities and challenges in relation to curriculum development, pedagogy, and student engagement [4]. Additional studies show that digital technologies and approaches can efficiently improve design education, thus equipping students to work in a more digital, mobile job industry [5]. The other prominent point of this development relates to the changing pedagogical behavior of teachers and the increased interest in digital means of teaching and learning. This pattern shows how traditional design teaching methods are evolving with the rise in digital technologies, with research examining their potential future effects on design education and the broader professional landscape [6]. A review of the modern tendencies in design education also highlights the influence of software, digital fabrication, and collaborative tools on educators and learners as digital tools start becoming part of design practice [7,8]. Though these global observations exist, there is still a lack of research, especially of a regional nature, particularly in the case of Jordan. Jordanian universities, which are progressively prioritizing integration of technology in education, have not comprehensively explored the impact of digital transformation on design education in the country. The current paper focused on filling this gap by performing a survey of teachers at Jordanian universities in a bid to understand their perceptions, challenges, and plans in terms of integrating digital tools in design education.

This paper uses three established theoretical frameworks to inform the analysis and provide a more in-depth theoretical insight. The degree of technological integration is assessed through the use of the SAMR model [9], which is a continuum of tools that started with simple substitution and ended with transformational change in education. These tools create cognitive levels ranging from basic knowledge retrieval to creative design, and are used to evaluate the level of cognitive engagement created by each tool using Bloom's Digital Taxonomy [10]. Lastly, the Technology Acceptance Model (TAM) [11] is utilized to examine the factors that determine the adoption of these tools with specific reference to the sense of ease of use and usefulness. The adoption of those theoretical models grounds this study in relevance to Human–Computer Interaction (HCI) and digital pedagogy, and through the implementation of these frameworks—crucial for understanding how digital tools are not merely adopted but integrated into pedagogical practices—it examines their role in supporting active learning, user engagement, and creativity, core tenets of HCI and modern educational technology, while offering a detailed report on the usage of digital tools in design learning, which also illustrates a descriptive and critical overview of the current situation of digital transformation in Jordan.

2. Literature Review

Over the past few years, there has been a growing appeal for using digital technologies in design education, partly because of the need for both educators and learners to stay abreast with an increasingly dynamic educational environment [12]. Online technologies like design software and group work platforms have transformed teaching and practice of design and created new possibilities and problems in education and learning [13]. An increasing number of studies are investigating these changes as a valuable source of information on how digital technologies may either support or disrupt learning [5,14,15].

The use of digital tools has significantly influenced the field of design education by making the process of learning more flexible and interactive [16]. The impact of digital technologies on the field of design education is proven in the studies, which affirm that the tools create a more adaptable and flexible environment of learning [17,18]. Researchers argue that digital technologies have the potential to promote different types of learning, as well as deliver instant feedback, and promote remote teamwork and individual creativity,

as well as group problem-solving skills [18]. However, it is recommended that they should be used sparingly: digital tools may be integrated to an extent that will lead to the stifling of critical design thinking [19].

Recent research also highlights the importance of interactive learning environments in fostering the engagement and creativity of students [20]. By using digital tools like software for 3D modeling and collaborative tools, students demonstrate improved interest in learning and experience active exchange during the design process. Also, they explore the digital tools that refine their creative output [21], which demonstrates the need to not only use digital tools but also make sure that the usage of those digital tools facilitates peer interaction and exploration.

Online resources have also been found to be effective in augmenting collaborative work, which is an essential part of design procedures [22]. Education using digital technologies enables collaborative learning, such as cloud-based tools and digital sketching tools that enable real-time feedback and collaboration, irrespective of geographic location [23]. This kind of partnership cannot be avoided in dealing with sophisticated design issues because different perspectives would lead to more innovative ideas. These technologies require the educators as well as the students to become skilled in online communication and teamwork.

Research into the adoption of digital design tools in architectural education within the fields of interior design, architecture, and urban design has compared the conventional learning environment to the digital learning environment with references to the use of computer-aided design software (CAD software), 3D modeling software, and technologies of digital fabrication, like 3D printing and computer numeric control machining [24,25]. Digital tools not only sped up and improved the design process, but they also gave students more freedom to be creative by letting them rapidly construct prototypes and try new things [26]. REVIT is mostly used in building information modeling (BIM)-based processes and is especially applicable in large, data-intensive architectural projects. The software also gives special emphasis on interdisciplinary coordination, thus making it very essential in producing construction documentation [27]. Rhino, in turn, is more focused on complex geometric lines and parametric design; hence, it can afford a greater degree of creative freedom. This is particularly relevant to areas requiring the development of freeform forms, like industrial design and architecture [28]. These technologies bring numerous advantages; nevertheless, they encounter some limitations, such as the high cost and difficulty of obtaining the right tools and the long process of learning that comes with understanding how to use complicated software.

Many studies regarding how digital fabrication changes the way people learn about design have been conducted. This study examined the potential of digital fabrication tools, such as 3D printers and CNC machines, in enhancing the process of design education [29]. The digital tools and technologies have significantly altered the process of prototyping, which enables students to swiftly create physical models of their designs, evaluate their concepts, and implement modifications with greater ease [30]. Students learn more about how things fit together in space and how materials work when they can quickly see and test design ideas in the real world. This bridges the gap between theoretical learning and practical application [31].

A significant progression in design education is the increasing application of virtual reality (VR) and immersive technologies, alongside their integration into design curricula, particularly in spatial design and architecture [32]. VR environments let students interact with their designs in very realistic ways, which helps them comprehend scale, space, and how components are related in space [33]. VR tools give students a chance to explore their designs in a virtual world [11], which makes learning more interactive and helps them

understand and solve problems better [34]. However, it is also noted that the high cost of VR technology and the need for specialized training can make it hard for many people to use it [35].

Along with the technological tools themselves, digital platforms that make it easier to learn and work together online are becoming more common in design education [36]. The role of online platforms in design education, especially in remote learning settings, is that they have made design education easier to access by letting students from different places work together on projects [37]. While digital platforms do offer more freedom, they also make it harder for students to connect and feel like they belong to a community [38].

Finally, digital narration is becoming an increasingly popular way to teach design [39], particularly through the use of digital visual narrative in teaching design, especially in architecture and interior design classes [40]. Using technology to communicate helps students share their design ideas and also improves their narrative abilities and creative skills. Students learn to think critically about how design elements and communication work together when they make digital stories [41]. This is an important skill in many design fields [42], where the success of digital narratives relies on instructors' capacity to navigate students through the process and offer constructive feedback [43,44].

These studies show that digital technologies have a complicated effect on design education. There are a lot of good things about using them, like more freedom, creativity, and teamwork. However, some problems need to be fixed. These include making sure that everyone has equal access to digital tools, establishing the right balance between creativity and technical skill, and figuring out how to manage digital collaboration well. As digital tools keep changing, we need to do more research to find out how they impact design education in the long run and how to use them in class best [45–48].

3. Materials and Methods

3.1. Research Design

This study employed a survey-based research design to examine the integration of digital tools in architecture and interior design education at Jordanian universities. The primary objective was to ascertain the utilization of these tools by educators, their proficiency with them, and their impact on students' creativity and engagement in learning. The methodological structure of the study is summarized in Figure 1.

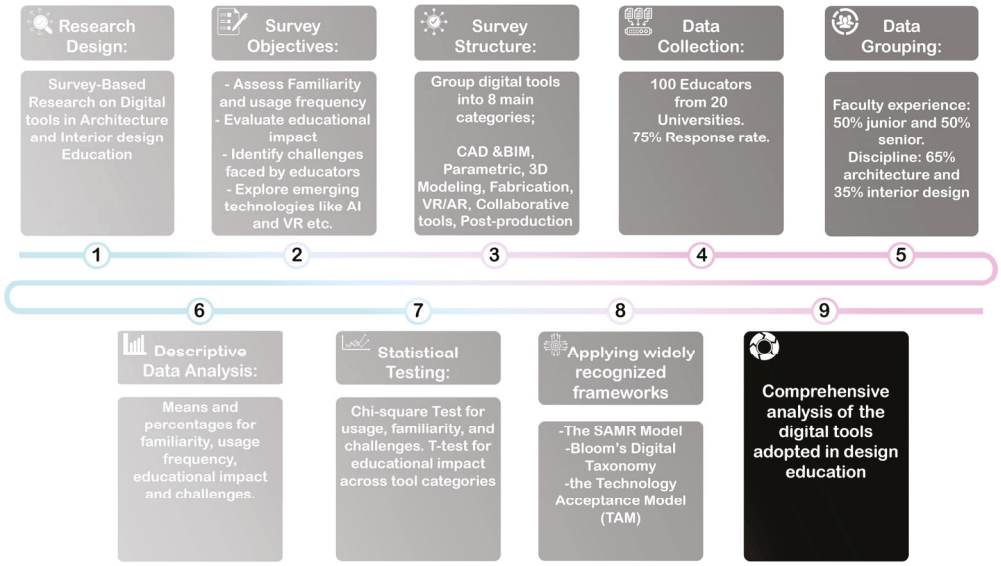


Figure 1. Methodological structure of the study (authors).

3.2. Survey Objectives

1. Evaluation of the familiarity and frequency of use of various digital tools in design education.
2. Determine how people think these tools will affect teaching and learning outcomes.
3. Identify challenges in integrating these tools into curricula.
4. Examine the role of emerging technologies (e.g., AI, VR) in design education.

3.3. Survey Structure and Tool Selection

The survey was structured around eight categories of digital tools commonly used in design education and as prevalent in relevant reviewed studies:

1. CAD and BIM Software: Tools for architectural design and modeling (e.g., AutoCAD, Revit, Rhino).
2. Parametric and Algorithmic Design Software: Tools for generative and algorithmic design (e.g., Grasshopper, Dynamo, Houdini).
3. 3D Modeling and Rendering Tools: Tools for visualizing and rendering design concepts (e.g., Lumion, V-Ray, 3ds Max).
4. Digital Fabrication and Prototyping Tools: Tools for creating physical prototypes (e.g., 3D printers, CNC machines).
5. Virtual and Augmented Reality Tools: Tools for immersive and interactive design (e.g., Unreal Engine, Unity).
6. Collaborative Design and Communication Tools: Tools for teamwork and project management (e.g., Miro, Figma, Google Drive).
7. Image Editing and Post-Production Software: Tools for editing and presenting design visuals (e.g., Photoshop, Illustrator).
8. AI Tools for Design: AI-driven tools for design generation and editing (e.g., Runway ML, DALL·E).

Even though eight categories of digital tools are assessed in this study, the wide range is not accidental. The target is to acquire an entire picture of the current adoption of various technologies, such as the CAD/BIM software, newly developed applications, such as VR/AR, and AI usage in the design education in Jordan. This approach to the methodology allows visualizing the entire picture of the digital transformation instead of focusing on the specific tools. The patterns of comparison that arise between these categories provide an understanding of the maturity of the adoption of digital and the trends in pedagogy that can be observed within design programs.

3.4. Survey Design and Implementation

The survey used closed-ended questions with a Likert scale to measure:

- Familiarity with the tools (Very Familiar, Somewhat Familiar, Not Familiar).
- Usage frequency (Daily, Weekly, Monthly, Rarely).
- Educational impact (Strongly Agree, Agree, Neutral, Disagree).
- Challenges (e.g., software cost, lack of training, limited access to computers).

A total of 100 educators from 20 universities (both private and public) offering architecture and interior design programs participated, resulting in a 75% response rate.

3.5. Data Analysis

Data were analyzed using descriptive statistics (means and percentages) to

- Identify the most commonly used tools across all categories.
- Assess familiarity with each tool.
- Evaluate the perceived educational impact of the tools.
- Identify the challenges educators face in implementing these tools in their teaching.

3.6. Statistical Analysis for Significance

To assess the significance of differences across tool categories, the following statistical tests were performed:

1. Chi-Square Test of Independence: This test was used to assess whether there were significant differences in familiarity, usage frequency, and challenges across the eight tool categories.
   - Null Hypothesis (H_0): There is no significant difference across the categories.
   - Alternative Hypothesis (H_1): There is a significant difference across the categories.
   - Significance Level: A p-value < 0.05 was considered statistically significant, indicating a significant difference between tool categories.
2. *T*-Test for Independent Samples: This test was used to compare the perceived educational impact between tool categories. It compared responses of "Strongly Agree" and "Agree" regarding educational impact, determining whether there were significant differences in perceptions of the tools' educational value.
   - Null Hypothesis (H_0): There is no significant difference in the educational impact between tool categories.
   - Alternative Hypothesis (H_1): There is a significant difference in the educational impact between tool categories.
   - Significance Level: A p-value < 0.05 was used to assess statistical significance. If the p-value was below 0.05, the null hypothesis was rejected, indicating a significant difference in perceived educational impact across tool categories.

3.7. Data Collection and Grouping

The respondents were grouped based on

- Faculty experience: 50% junior faculty (0–5 years) and 50% senior faculty (6+ years).
- Discipline: 65% architecture faculty and 35% interior design faculty.

Furthermore, the data used in this study were fully anonymized and did not contain identifiable human subject information; therefore, informed consent was not required.

3.8. Statistical Software

Data analysis was conducted using SPSS (Statistical Package for the Social Sciences, version 30). The Chi-Square and *T*-tests were chosen due to their ability to analyze categorical and continuous data, respectively. For both tests, the threshold for statistical significance was set at a p-value < 0.05.

3.9. Applying Widely Recognized Theoretical Frameworks

This study uses three theoretical frameworks, i.e., SAMR, Bloom's Digital Taxonomy, and Technology Acceptance Model (TAM), to understand the integration of digital tools in design education. A detailed description of the operationalization of each framework in the context of the present investigation follows below.

a. The SAMR model (Substitution, Augmentation, Modification, and Redefinition) will be used to evaluate the level of technological integration that was demonstrated by the digital tools that were used for design education.
   - Procedure: Survey data will be analyzed, and each digital tool will be classified by one of the four SAMR stages. This classification will be based on the educators' reports on the usage of the tools in their practices.

- Substitution: Tools that replace conventional methods but do not change the education experience much.
- Augmentation: Tools that offer improvements over traditional methods and, thus, functional enhancements that are added to the learning process.
- Modification: Tools that allow for a great deal of redesign of learning tasks, thus enabling more creative and complex design work.
- Redefinition: Tools that give rise to new tasks that are previously inconceivable and, thus, provide transformative learning experiences.

b. Bloom's Digital Taxonomy will be used to assess cognitive engagement promoted in each digital tool. The taxonomy outlines a hierarchy of cognitive skills from the lower-order to the higher-order thinking.

- Procedure: For each tool, data will be surveyed to determine which cognitive levels are supported the most. Tools will be evaluated based on the types of tasks that they will allow students to do.
- Lower-order cognitive tasks, such as Remembering and Applying, will be associated with tools that support basic design functions (e.g., AutoCAD).
- Higher-order cognitive tasks that will be linked to tools that foster creativity and increased engagement with the design process (e.g., VR/AR).

c. The Technology Acceptance Model will be used for the evaluation of perceived ease of use and perceived usefulness of each digital tool, as these perceptions impact the propensity to use the tools in educational settings.

- Procedure: Respondents in the survey will be asked to rate each tool in terms of perceived ease of use and perceived usefulness. These ratings will then be analyzed to see which attributes—ease of use, usefulness, or both—are most important for adoption.
- Perceived Ease of Use: How easy the educators feel the tool is to use, with little effort or special training required.
- Perceived Usefulness: How much educators perceived that the use of the tool increased the effectiveness of teaching and student learning outcomes [9–11].

4. Results

The survey was completed by 100 respondents from 20 universities (both private and governmental) offering architecture and interior design programs. The data provides insights into how these tools are perceived and used in design education in Jordan. Taking into account that the number of universities contributing is 20 (13 with Architecture departments, 7 with Interior Design departments, and some universities offering both), the number of respondents per university was 5 respondents, bringing the total respondents participating in the study from the 20 universities to 100 respondents.

4.1. Data Analysis Results

The descriptive analysis Results Section is presented with the following structure for each of the eight categories:

A. Familiarity: Present the familiarity data, followed by the chart and its interpretation.
B. Usage Frequency: Present the usage frequency data, followed by the chart and its interpretation.
C. Educational Impact: Present the educational impact data, followed by the chart and its interpretation.
D. Challenges: Present the challenges data, followed by the chart and its interpretation.

4.1.1. Results Section for the First Category; CAD/BIM Tools

A. Familiarity with CAD/BIM Tools, as shown in Table 1.

Table 1. Familiarity with CAD/BIM tools (authors).

Tool	Very Familiar (%)	Somewhat Familiar (%)	Not Familiar (%)
AutoCAD	80	15	5
Revit	65	25	10
Rhino	40	35	25
Archicad	55	30	15
SketchUp	75	20	5

Results from this analysis show that AutoCAD and SketchUp have the highest familiarity, with 80% and 75% of respondents being very familiar with them, respectively. This indicates that these tools are well-established and commonly used in design education. While Revit and Archicad have lower familiarity levels (65% and 55%), suggesting that these tools are important but less universally recognized or adopted compared to AutoCAD. And finally, Rhino shows the lowest level of familiarity, with only 40% of respondents reporting high familiarity, possibly reflecting its more specialized use.

B. Usage Frequency of CAD/BIM Tools, as shown in Table 2.

Table 2. Usage frequency of CAD/BIM tools (authors).

Tool	Daily (%)	Weekly (%)	Monthly (%)	Rarely (%)
AutoCAD	50	30	15	5
Revit	25	40	25	10
Rhino	10	25	40	25
Archicad	30	35	20	15
SketchUp	60	25	10	5

Results from this analysis show that SketchUp has the highest daily usage (60%), followed by AutoCAD (50%). These tools are used most frequently, suggesting their importance in daily design tasks. Revit shows moderate usage, with 25% of respondents using it daily and 40% using it weekly. This reflects its importance but indicates it may not be integrated into daily teaching. Meanwhile, Rhino and Archicad are used less frequently, with only 10% of respondents using Rhino daily. These tools are likely used for more specialized tasks and may require additional training or resources.

C. Educational Impact of CAD/BIM Tools, as shown in Table 3.

Table 3. Educational impact of CAD/BIM tools (authors).

Tool	Strongly Agree (%)	Agree (%)	Neutral (%)	Disagree (%)
AutoCAD	70	20	10	0
Revit	40	35	15	10
Rhino	20	30	40	10
Archicad	50	30	15	5
SketchUp	80	15	5	0

Results from this analysis show that SketchUp has the highest positive impact, with 80% of educators strongly agreeing that it enhances student learning. Its ease of use and visualization capabilities make it an essential tool in design education. AutoCAD also

shows a strong impact, with 70% strongly agreeing that it improves teaching and learning. Revit has a more mixed impact, with only 40% strongly agreeing. Given its complexity, educators may need more training to integrate it into teaching fully. And finally, Rhino shows a lower educational impact, with only 20% strongly agreeing. This could be due to its specialized nature and higher learning curve.

D. Challenges in Using CAD/BIM Tools, as shown in Table 4.

Table 4. Challenges in using CAD/BIM tools (authors).

Challenge	AutoCAD (%)	Revit (%)	Rhino (%)	Archicad (%)	SketchUp (%)
High Software Cost	60	60	60	50	40
Lack of Training or Expertise	55	55	50	60	35
Limited Access to Computers	45	45	45	40	30
Technical Difficulties	30	30	35	25	20

Results from this analysis show that the most significant challenge reported for all tools is the high software cost (60%). This reflects the expensive licensing for professional design tools, which may limit access in educational institutions. Lack of training and limited access to computers were also significant barriers, especially for tools like Revit and Rhino that require specialized hardware and expertise. Technical difficulties were less common but still notable for tools like Rhino (35%) and Revit (30%), possibly indicating the complexity of these tools.

The following are the charts corresponding to the descriptive analysis for the Results Section for the first category, CAD/BIM tools, as shown in Figure 2.

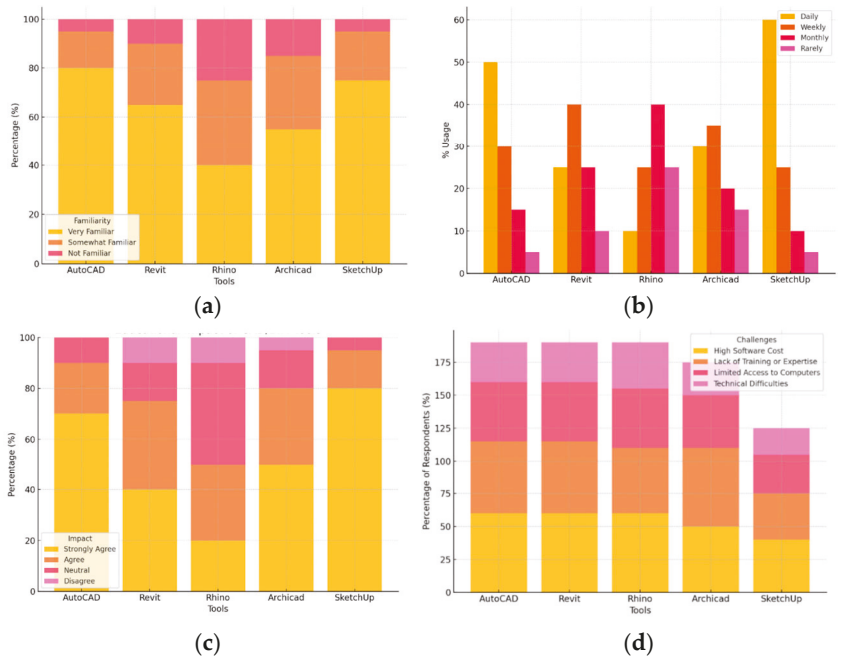


Figure 2. (**a**): Familiarity, (**b**): usage frequency, (**c**): educational impact, and (**d**): challenges in using CAD/BIM tools (authors).

4.1.2. Results Section for the Second Category; Parametric and Algorithmic Design Software

A. Familiarity with Parametric and Algorithmic Design Software, as shown in Table 5.

Table 5. Familiarity with Parametric and Algorithmic Design Software (authors).

Tool	Very Familiar (%)	Somewhat Familiar (%)	Not Familiar (%)
Grasshopper (Rhino)	70	20	10
Dynamo (Revit)	50	35	15
Houdini	30	40	30
TouchDesigner	20	50	30

Results from this analysis show that Grasshopper (for Rhino) has the highest familiarity, with 70% of respondents being very familiar with it. This is consistent with its widespread use in architectural and design education, especially in programs that focus on parametric design. Dynamo (for Revit) shows a moderate level of familiarity, with 50% of respondents reporting that they are very familiar with it. Dynamo is used in BIM workflows but is not as universally known as Grasshopper, reflecting its specialized application. Houdini and TouchDesigner have lower familiarity levels, with 30% and 20% of respondents being very familiar with them, respectively. These tools are more specialized for advanced design and animation, which may explain their lower adoption rate in standard design curricula.

B. Usage Frequency of Parametric and Algorithmic Design Software, as shown in Table 6.

Table 6. Usage frequency of Parametric and Algorithmic Design Software (authors).

Tool	Daily (%)	Weekly (%)	Monthly (%)	Rarely (%)
Grasshopper (Rhino)	40	30	20	10
Dynamo (Revit)	20	40	30	10
Houdini	10	30	40	20
TouchDesigner	5	20	50	25

Results from this analysis show that Grasshopper is the most commonly used tool on a daily basis, with 40% of respondents using it regularly. This highlights Grasshopper's importance in design education, particularly in courses focused on parametric design and computational design methods. Dynamo is used more moderately, with 20% of respondents using it daily. This suggests that while Dynamo is important in BIM environments, it may not be as frequently used as Grasshopper in non-BIM-focused design curricula. Houdini and TouchDesigner have much lower daily usage, with 10% and 5% of respondents using them daily, respectively. These tools are likely reserved for more advanced, specialized design courses and are not yet integrated into the core curriculum for most design programs.

C. Educational Impact of Parametric and Algorithmic Design Software, as shown in Table 7.

Table 7. Educational impact of Parametric and Algorithmic Design Software (authors).

Tool	Strongly Agree (%)	Agree (%)	Neutral (%)	Disagree (%)
Grasshopper (Rhino)	60	30	10	0
Dynamo (Revit)	40	40	15	5
Houdini	25	40	25	10
TouchDesigner	15	45	30	10

Results from this analysis show that Grasshopper shows the strongest educational impact, with 60% of respondents strongly agreeing that it enhances teaching and learning.

Its integration with Rhino allows for rapid design exploration and facilitates computational thinking, making it a powerful educational tool in design studios. Dynamo also shows a positive impact, with 40% of respondents strongly agreeing that it improves education. As a tool for integrating design and BIM workflows, it has a strong educational impact, particularly for students learning BIM processes. Houdini and TouchDesigner have lower perceived educational impact, with only 25% and 15% strongly agreeing, respectively. While these tools are powerful for 3D modeling and animation, they are more specialized and may not have widespread educational application in general design programs.

D. Challenges in Using Parametric and Algorithmic Design Software, as shown in Table 8.

Table 8. Challenges in using Parametric and Algorithmic Design Software (authors).

Challenge	Grasshopper (%)	Dynamo (%)	Houdini (%)	TouchDesigner (%)
High Software Cost	60	60	55	50
Lack of Training	55	55	60	50
Limited Access to Computers	45	45	40	30
Technical Difficulties	40	40	45	35

Results from this analysis shows that high software cost is the most common challenge, reported by 60% of respondents across all tools. This highlights the barrier that the high cost of professional-grade software (like Rhino, Revit, Houdini, etc.) can have in limiting access to these tools, especially in institutions with constrained budgets. Lack of training is the second most common challenge, with 55% of respondents citing it, particularly for more advanced tools like Houdini and TouchDesigner. This indicates that even when tools are available, educators may lack the resources or expertise to teach them effectively. Limited access to computers is also an important challenge, reported by 45% of respondents, reflecting the fact that some of these tools require significant computational power, which may not be available in all institutions. Technical difficulties are reported by 40% of respondents, particularly for TouchDesigner and Houdini, which may be more complex to use and integrate into educational curricula.

The following are the charts corresponding to the descriptive analysis for the Results Section for the Second Category, Parametric and Algorithmic Design Software, as shown in Figure 3.

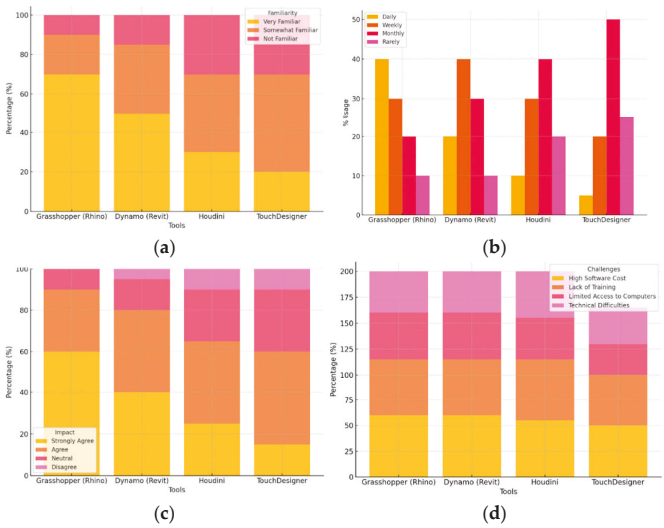


Figure 3. (**a**): Familiarity, (**b**): usage frequency, (**c**): educational impact, and (**d**): challenges in using Parametric and Algorithmic Design Software (authors).

4.1.3. Results Section for the Third Category; 3D Modeling and Rendering Tools

A. Familiarity with 3D Modeling and Rendering Tools, as shown in Table 9.

Table 9. Familiarity with 3D Modeling and Rendering Tools (authors).

Tool	Very Familiar %	Somewhat Familiar %	Not Familiar %
Lumion	75	20	5
Twinmotion	70	25	5
3ds Max	65	30	5
V-Ray	80	15	5
Corona Renderer	60	35	5

Results from this analysis show that V-Ray is the most familiar tool, with 80% of respondents reporting being very familiar with it. This reflects its widespread use in design education and industry for rendering purposes. Lumion and Twinmotion also have high familiarity rates, with 75% and 70% of respondents being very familiar, respectively. These tools are popular for their user-friendly interfaces and real-time rendering capabilities. 3ds Max shows moderate familiarity, with 65% reporting a high level of familiarity. While it is widely used in professional design settings, it may be less commonly incorporated into design education compared to tools like V-Ray. Corona Renderer has the lowest familiarity at 60%, though still relatively high. Its use in educational settings may be more recent or less prevalent than the other tools.

B. Usage Frequency of 3D Modeling and Rendering Tools, as shown in Table 10.

Table 10. Usage frequency of 3D Modeling and Rendering Tools (authors).

Tool	Daily (%)	Weekly (%)	Monthly (%)	Rarely (%)
Lumion	40	35	20	5
Twinmotion	35	40	20	5
3ds Max	25	45	25	5
V-Ray	30	40	20	10
Corona Renderer	20	50	25	5

The analysis shows that Lumion and Twinmotion have the highest percentage of daily usage, with 40 percent and 35 percent of the respondents, respectively, indicating that they use the tools daily. They are effective because of their fast-processing speeds and user-friendly interfaces, which ensure that realistic visualizations are produced. Three-de-eight-seo (3ds 3Ds Max) shows moderate usage or 25 percent of respondents; its relatively complex nature could restrict its ease of use in everyday classroom use by some individuals. The frequency of V-Ray and Corona Renderer is utilized at a lower rate on a daily basis, with the daily use reported at 30, indicating that people use it daily, and 20 say that they use it daily. These are also highly specialized applications, with the workflow of producing high-quality rendering often requiring more time and computation capacity to produce detailed visualization.

C. Educational Impact of 3D Modeling and Rendering Tools, as shown in Table 11.

Table 11. Educational impact of 3D Modelling and Rendering Tools (authors).

Tool	Strongly Agree (%)	Agree (%)	Neutral (%)	Disagree (%)
Lumion	70	25	5	0
Twinmotion	65	30	5	0

Table 11. *Cont.*

Tool	Strongly Agree (%)	Agree (%)	Neutral (%)	Disagree (%)
3ds Max	50	40	10	0
V-Ray	75	15	10	0
Corona Renderer	60	35	5	0

The same analytical framework reveals that V-Ray has an outstanding educational presence, with 75 percent of the respondents providing a high recommendation of effectiveness in the promotion of teaching and learning. The high fidelity and versatility of the tool make it a great asset in the design pedagogy. Lumion and Twinmotion also exhibit a high positive impact, with 70 percent and 65 percent of teachers confirming that they contribute to student learning, respectively. Their ability to provide real-time visualizations is particularly useful in enabling them to perform quick design iterations. Three-de-e-eight-seo is moderately positive; 50 percent of the respondents claim to agree with it strongly, but its complexity might require further training and practice, yet it provides advanced functionality to build complex 3D constructions. Corona Renderer has slightly less impact when compared to V-Ray, but is still positive, where 60 percent strongly agreed that it maximizes educational output. The high-quality design communication is supported by the photorealism of the tool.

D. Challenges in Using 3D Modeling and Rendering Tools, as shown in Table 12.

Table 12. Challenges in using 3D Modeling and Rendering Tools (authors).

Challenge	Lumion %	Twinmotion %	3ds Max %	V-Ray %	Corona Renderer %
High Software Cost	50	45	60	55	50
Lack of Training	40	40	45	35	45
Limited Access to Computers	30	35	40	30	35
Technical Difficulties	25	30	35	40	30

Results from this analysis show that high software cost is the most significant challenge across all tools, with 60% of respondents reporting this challenge for 3ds Max. The high cost of professional software licenses is a common barrier to their widespread use in educational settings. Lack of training is also a notable challenge, particularly for tools like 3ds Max and Corona Renderer, where 45% of respondents report this barrier. This highlights the need for professional development and training to integrate these tools into curricula effectively. Limited access to computers is a challenge, especially for tools that require high-performance hardware, such as 3ds Max and V-Ray. Technical difficulties were less common but still notable, particularly for V-Ray, where 40% of educators reported challenges in handling the tool's complex settings. The following are the charts corresponding to the descriptive analysis for the Results Section for the Third Category: 3D Modeling and Rendering Tools, as shown in the charts in Figure 4.

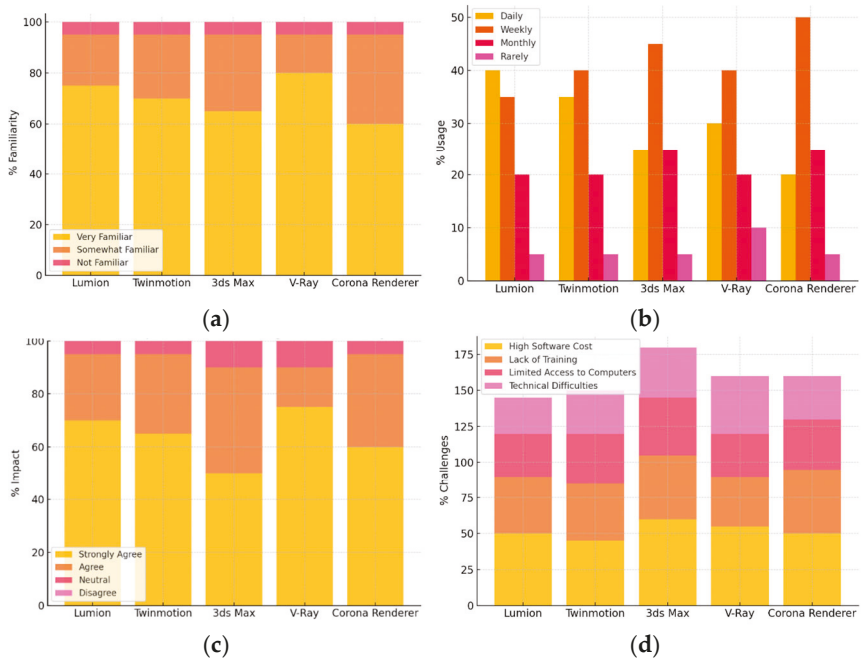


Figure 4. (**a**): Familiarity, (**b**): usage frequency, (**c**): educational impact, and (**d**): challenges in using 3D Modeling and Rendering Tools (authors).

4.1.4. Results Section for the Fourth Category; Digital Fabrication and Prototyping Tools

A. Familiarity with Digital Fabrication and Prototyping Tools, as shown in Table 13.

Table 13. Familiarity with Digital Fabrication and Prototyping Tools (authors).

Tool	Very Familiar %	Somewhat Familiar %	Not Familiar %
Laser Cutters	60	30	10
CNC Milling Machines	50	35	15
3D Printers	75	15	10
Router Machines	65	25	10

Regarding familiarity, the 3D printers have the most recognition, with 75% of respondents indicating high familiarity. This tool is widely used in academic and professional environments to prototype quickly and is becoming a part of design courses. The moderate familiarity is reached by laser cutters and router machines, which have the reported levels of 60 and 65%, respectively; these tools are predominantly used in the production of precise and detailed models, albeit with specialized knowledge. CNC milling machines are the least familiar, with 50% of the participants reporting a high familiarity. Although they are common in industrial and architectural fabrication, they are also not as commonly used in design education programs as 3D printers.

B. Usage Frequency of Digital Fabrication and Prototyping Tools, as shown in Table 14.

Table 14. Usage frequency of Digital Fabrication and Prototyping Tools (authors).

Tool	Daily (%)	Weekly (%)	Monthly (%)	Rarely (%)
Laser Cutters	40	30	20	10
CNC Milling Machines	35	30	25	10
3D Printers	50	30	15	5
Router Machines	45	25	20	10

Daily use analysis shows that 3D printers are the most used, with 50% of the respondents reporting daily usage. This trend indicates increased accessibility and teaching integration of 3D printing in design education for rapid prototyping. Laser cutters and router machines also demonstrate moderate usage with 40% and 45 percent daily use, respectively; these tools are necessary to make precise fabrication but are not used as often as 3D printers. CNC milling machines show the lowest usage per day, with 35 percent of the participants using the machines weekly, which indicates that their application is more associated with sophisticated fabrication operations than everyday teaching.

C. Educational Impact of Digital Fabrication and Prototyping Tools, as shown in Table 15.

Table 15. Educational impact of Digital Fabrication and Prototyping Tools (authors).

Tool	Strongly Agree %	Agree %	Neutral %	Disagree %
Laser Cutters	75	20	5	0
CNC Milling Machines	60	30	10	0
3D Printers	85	10	5	0
Router Machines	70	20	10	0

The last evaluation of educational impact signifies that the role of 3D printers has the strongest impact, with 85% of the respondents strongly supporting this as a tool to enhance teaching and learning. This was influenced by their versatility and continued spread to design education. Laser cutters and router machines also exhibit a positive effect, with 75% and 70% of respondents saying that the tools have improved learning outcomes, respectively. These tools are essential for accuracy work and may significantly enliven practical designing experiences. The influence of CNC milling machines is somewhat more uneven, with 60% of the respondents strongly agreeing that the machines might help in enhancing education, reflecting a more specialized use in the wider education spectrum.

D. Challenges in Using Digital Fabrication and Prototyping Tools, as shown in Table 16.

Table 16. Challenges in using Digital Fabrication and Prototyping Tools(authors).

Challenge	Laser Cutters %	CNC Milling Machines %	3D Printers %	Router Machines %
High Software Cost	60	55	50	50
Lack of Training	55	50	45	50
Limited Access to Computers	45	40	35	40
Technical Difficulties	40	45	35	30

The results show that high software cost is the main bottleneck in all digital fabrication tools, with laser cutters (60%) and CNC milling machines (55%) being the most vulnerable. Next in line is the lack of training, especially among laser cutters (55%) and CNC milling machines (50%), which demonstrates the need to have specific knowledge. Limited access to computing facilities affects all devices, but with a significant effect on 3D printers (35%), this further intensifies the need to have high-performance hardware. Although the technical complications are less common, they are still consequential, especially the router machines (30%), with the lowest rate of occurrence. The leading barriers that hinder the successful implementation of these technologies into design pedagogy include cost, training, and availability of resources, collectively.

The following are the charts corresponding to the descriptive analysis for the Results Section for the Fourth Category, Digital Fabrication and Prototyping Tools, as shown in Figure 5.

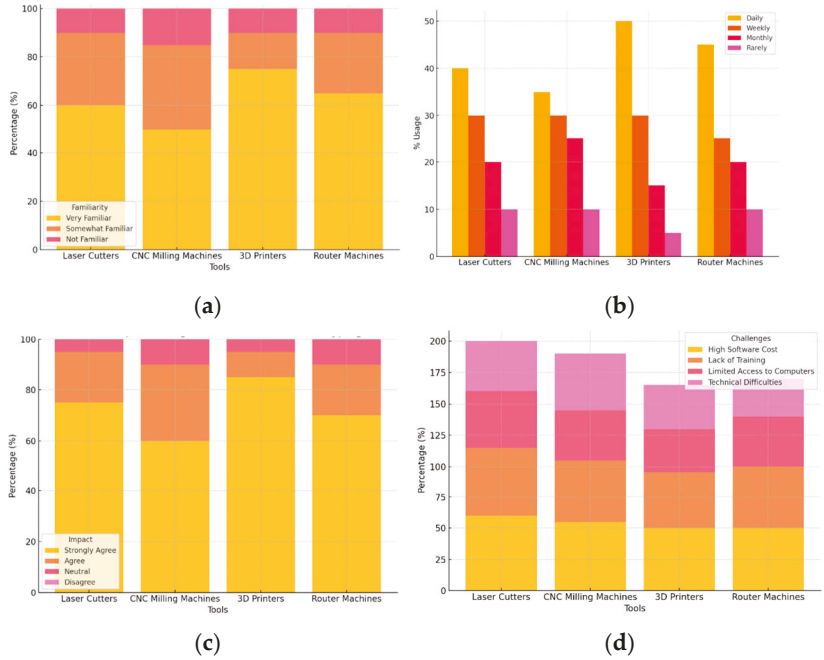


Figure 5. (**a**): Familiarity, (**b**): usage frequency, (**c**): educational impact, and (**d**): challenges in using Digital Fabrication and Prototyping Tools (authors).

4.1.5. Results Section for the Fifth Category; Virtual and Augmented Reality Tools

A. Familiarity with Virtual and Augmented Reality Tools, as shown in Table 17.

Table 17. Familiarity with Virtual and Augmented Reality Tools (authors).

Tool	Very Familiar %	Somewhat Familiar %	Not Familiar %
Unreal Engine	60	30	10
Unity	55	35	10
Enscape	70	20	10
IrisVR	40	40	20
Hololens	30	30	40

Results from this analysis show that Enscape is the most familiar tool, with 70% of respondents being very familiar with it. It is widely used in design education for immersive visualizations. Unreal Engine and Unity also have a strong familiarity rate of 60% and 55%, respectively. These tools are well-known for their interactive and immersive capabilities in design education. IrisVR and Hololens have lower familiarity levels, with 40% and 30% of respondents being very familiar. These tools are more specialized and may not yet be integrated into the core curricula of many design programs.

B. Usage Frequency of Virtual and Augmented Reality Tools, as shown in Table 18.

Table 18. Usage frequency of Virtual and Augmented Reality Tools (authors).

Tool	Daily (%)	Weekly (%)	Monthly (%)	Rarely (%)
Unreal Engine	40	35	20	5
Unity	35	40	20	5

Table 18. *Cont.*

Tool	Daily (%)	Weekly (%)	Monthly (%)	Rarely (%)
Enscape	45	35	15	5
IrisVR	25	45	25	5
Hololens	20	40	30	10

Results from this analysis show that Enscape has the highest daily usage at 45%, followed by Unreal Engine and Unity at 40% and 35%, respectively. These tools are frequently used in design studios for immersive visualizations. IrisVR and Hololens are used less frequently, with 25% and 20% using them daily, indicating that these tools might be more specialized or used in specific courses. IrisVR and Hololens show relatively high weekly usage, with 45% and 40% using them regularly, indicating that they are still being integrated into the curriculum but with moderate frequency.

C. Educational Impact of Virtual and Augmented Reality Tools, as shown in Table 19.

Table 19. Educational Impact of Virtual and Augmented Reality Tools (authors).

Tool	Strongly Agree (%)	Agree (%)	Neutral (%)	Disagree (%)
Unreal Engine	50	30	15	5
Unity	45	40	10	5
Enscape	60	25	10	5
IrisVR	35	40	20	5
Hololens	30	35	25	10

Results from this analysis show that Enscape has the highest positive impact, with 60% of educators strongly agreeing that it enhances teaching and learning. Its real-time rendering capabilities make it a powerful tool for design education. Unreal Engine and Unity also show strong educational impact, with 50% and 45% of educators strongly agreeing. These tools support immersive and interactive learning experiences. IrisVR and Hololens show a more mixed impact, with 35% and 30% strongly agreeing. While these tools are impactful for certain educational contexts, their specialized nature might limit their widespread application.

D. Challenges in Using Virtual and Augmented Reality Tools, as shown in Table 20.

Table 20. Challenges in Using Virtual and Augmented Reality Tools (authors).

Challenge	Unreal Engine (%)	Unity (%)	Enscape (%)	IrisVR (%)	Hololens (%)
High Software Cost	55	50	45	60	65
Lack of Training	45	40	50	45	40
Limited Access to Computers	40	35	30	40	50
Technical Difficulties	35	40	35	30	25

Results from this analysis show that high software cost is the most significant challenge across all tools, with 65% of respondents citing it for Hololens and 60% for IrisVR. The high cost of VR/AR hardware and software licenses is a common barrier in educational settings. Lack of training is also a major issue, with 50% of respondents citing it for Enscape. This suggests a need for more professional development and resources to fully integrate these tools into design education. Limited access to computers is a significant challenge, particularly for Hololens, with 50% of respondents indicating this as a barrier. VR/AR

tools often require high-performance computers or specialized hardware, which may not be available in all institutions. Technical difficulties were reported by 35% of respondents for Unreal Engine and Unity, indicating that while these tools offer great potential, they may require more technical expertise to integrate into the curriculum effectively.

The following are the charts corresponding to the descriptive analysis for the Results Section for the Fifth Category, Virtual and Augmented Reality Tools, as shown in Figure 6.

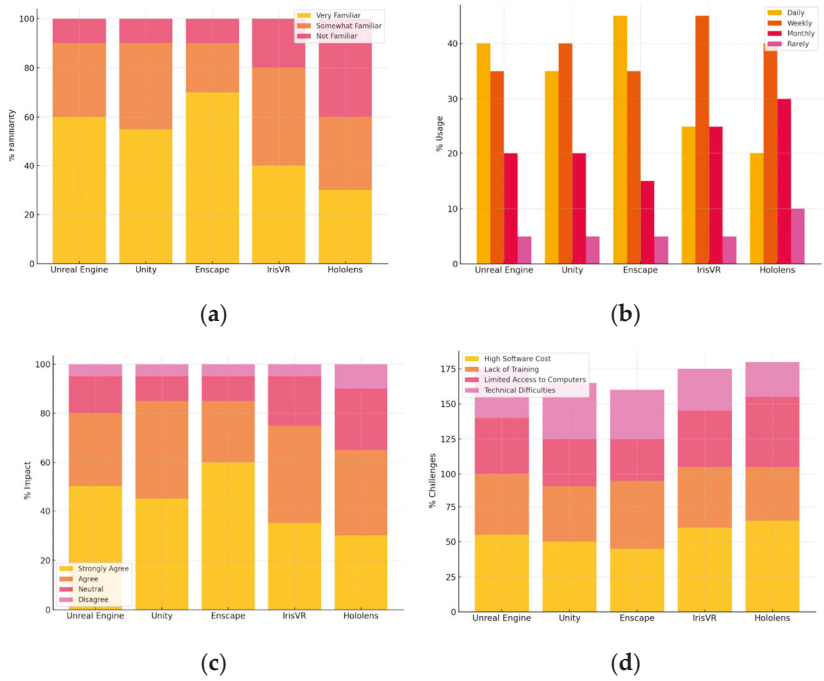


Figure 6. (**a**): Familiarity, (**b**): usage frequency, (**c**): educational impact, and (**d**): challenges in using Virtual and Augmented Reality Tools (authors).

4.1.6. Results Section for the Sixth Category; Image Editing and Post-Production Tools

A. Familiarity with Image Editing and Post-Production Software, as shown in Table 21.

Table 21. Familiarity with Image Editing and Post-Production Software (authors).

Tool	Very Familiar %	Somewhat Familiar %	Not Familiar %
Adobe Photoshop	85	10	5
Adobe Illustrator	80	15	5
CorelDRAW	60	20	20
GIMP	50	30	20
InDesign	70	15	15

Results from this analysis show that Adobe Photoshop is the most familiar tool, with 85% of respondents being very familiar with it. This reflects its widespread use in both professional design and educational settings for editing and enhancing visual content. Adobe Illustrator follows closely, with 80% of respondents being very familiar, confirming its significant role in vector graphic design. InDesign also shows high familiarity at 70%, as it is essential for layout design and final presentation work. CorelDRAW and GIMP have lower familiarity rates, with 60% and 50%, respectively, but still indicate solid usage in certain educational contexts, particularly in open-source environments (GIMP) and in traditional design settings (CorelDRAW).

B. Usage frequency of Image Editing and Post-Production Software, as shown in Table 22.

Table 22. Usage frequency of Image Editing and Post-Production Software (authors).

Tool	Daily (%)	Weekly (%)	Monthly (%)	Rarely (%)
Adobe Photoshop	70	25	5	0
Adobe Illustrator	60	30	5	5
CorelDRAW	25	40	20	15
GIMP	15	40	30	15
InDesign	50	35	10	5

The results of this survey show that 70% of the people who answered use Adobe Photoshop every day. This shows that it is an important part of design education for things like editing images, changing the way they look, and making them look better. Adobe Illustrator is also used a lot every day, at 60%, which makes sense because it is important for making designs and illustrations. Half of the people who answered said they use InDesign every day, which shows how important it is for layout and presentation design. CorelDRAW and GIMP are used less often every day, with 25% and 15% of users, respectively. However, both tools are still used a lot, especially in some specialized design courses or for open-source projects.

C. Educational Impact of Image Editing and Post-Production Software, as shown in Table 23.

Table 23. Educational Impact of Image Editing and Post-Production Software (authors).

Tool	Strongly Agree (%)	Agree (%)	Neutral (%)	Disagree (%)
Adobe Photoshop	85	10	5	0
Adobe Illustrator	80	15	5	0
CorelDRAW	60	30	10	0
GIMP	55	35	10	0
InDesign	75	20	5	0

Results from this analysis show that Adobe Photoshop has the highest educational impact, with 85% of educators strongly agreeing that it enhances teaching and learning. Its role in image manipulation, retouching, and creating design assets makes it indispensable in design education. Adobe Illustrator follows closely, with 80% of respondents strongly agreeing about its positive impact, highlighting its effectiveness in vector-based graphic design. InDesign also has a strong educational impact, with 75% of respondents strongly agreeing that it helps in layout design and final presentation work. CorelDRAW and GIMP show moderate educational impact, with 60% and 55% strongly agreeing, indicating their relevance in some design curricula, particularly in more traditional or open-source contexts.

D. Challenges in Using Image Editing and Post-Production Software, as shown in Table 24.

Table 24. Challenges in Using Image Editing and Post-Production Software (authors).

Challenge	Adobe Photoshop %	Adobe Illustrator %	CorelDRAW %	GIMP %	InDesign %
High Software Cost	60	55	50	40	45
Lack of Training	55	50	45	50	40
Limited Access to Computers	40	35	40	35	30
Technical Difficulties	25	30	35	30	20

Results from this analysis show that high software cost is the most significant challenge for all tools, particularly for Adobe Photoshop and Illustrator, with 60% and 55% of respondents reporting it. The high cost of licensing for these industry-standard tools can be a barrier in educational environments, especially in institutions with limited budgets. Lack of training is a notable issue, reported by 55% of respondents for Photoshop and 50% for Illustrator, indicating that educators may need more professional development or resources to teach these tools effectively. Limited access to computers is a challenge for all tools, particularly for Photoshop, Illustrator, and CorelDRAW, where 40% to 60% of respondents reported this issue. These tools often require high-performance computers, which may not be available in all institutions. Technical difficulties were less common but still notable, especially for CorelDRAW and GIMP, where 35% and 30% of respondents cited this barrier. These issues may be related to compatibility or software setup.

The following are the charts corresponding to the descriptive analysis for the Results Section for the Sixth Category, Image Editing and Post-Production Tools, as shown in Figure 7.

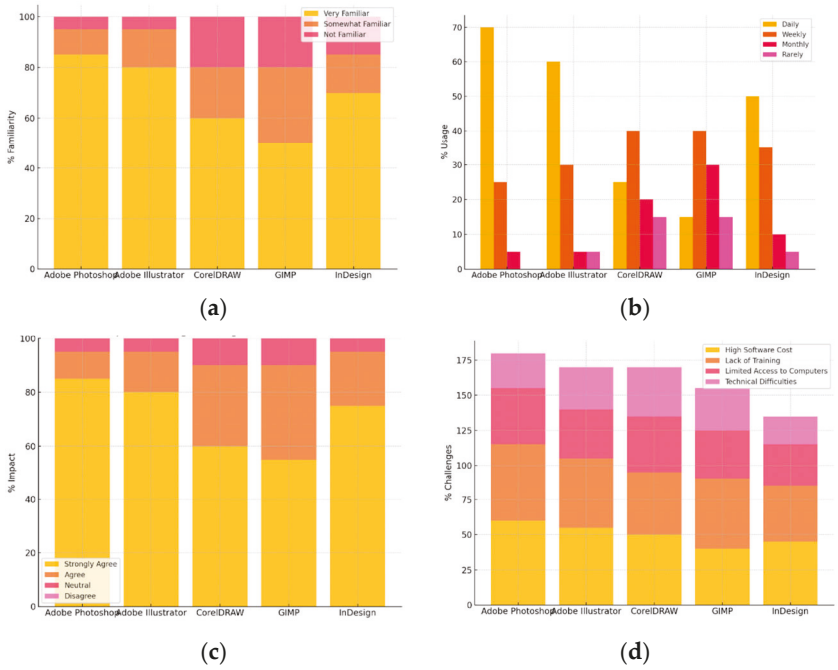


Figure 7. (**a**): Familiarity, (**b**): Usage Frequency, (**c**): Educational Impact, and (**d**): Challenges of Image Editing and Post-Production Tools (authors).

4.1.7. Results Section for the Seventh Category; Collaborative Design and Communication Tools

A. Familiarity with Collaborative Design and Communication Tools, as shown in Table 25.

Table 25. Familiarity with Collaborative Design and Communication Tools (authors).

Tool	Very Familiar (%)	Somewhat Familiar (%)	Not Familiar (%)
Miro	65	25	10
Figma	60	30	10
Trello	70	20	10
Slack	55	35	10
Google Drive	80	15	5

The familiarity data show a relatively high recognition of Google Drive (80%) and Trello (70%) as well-established tools in design education. In contrast, tools like Figma and Slack are moderately familiar but not as universally used.

B. Usage frequency of Collaborative Design and Communication Tools, as shown in Table 26.

Table 26. Usage frequency of Collaborative Design and Communication Tools (authors).

Tool	Daily (%)	Weekly (%)	Monthly (%)	Rarely (%)
Miro	35	40	20	5
Figma	30	50	15	5
Trello	50	35	10	5
Slack	40	45	10	5
Google Drive	60	30	10	0

The usage frequency data indicate that Google Drive is used most frequently (60% daily), followed by Trello (50%). However, tools like Miro and Figma are used less frequently.

C. Educational Impact of Collaborative Design and Communication Tools, as shown in Table 27.

Table 27. Educational Impact of Collaborative Design and Communication Tools (authors).

Tool	Strongly Agree (%)	Agree (%)	Neutral (%)	Disagree (%)
Miro	60	35	5	0
Figma	55	40	5	0
Trello	65	30	5	0
Slack	50	40	10	0
Google Drive	75	20	5	0

The educational impact data show a positive effect across all tools, with Google Drive and Trello leading in strong agreement for enhancing teaching and learning.

D. Challenges in Using Collaborative Design and Communication Tools, as shown in Table 28.

Table 28. Challenges in Using Collaborative Design and Communication Tools (authors).

Challenge	Miro (%)	Figma (%)	Trello (%)	Slack (%)	Google Drive (%)
High Software Cost	30	40	30	25	20
Lack of Training	40	35	40	30	25
Limited Access to Computers	25	20	25	20	15
Technical Difficulties	20	25	20	15	10

The most common challenge faced across all tools is a lack of training, followed by high software costs. The following are the charts corresponding to the descriptive analysis for the Results Section for the Seventh Category, Collaborative Design and Communication Tools, as shown in Figure 8.

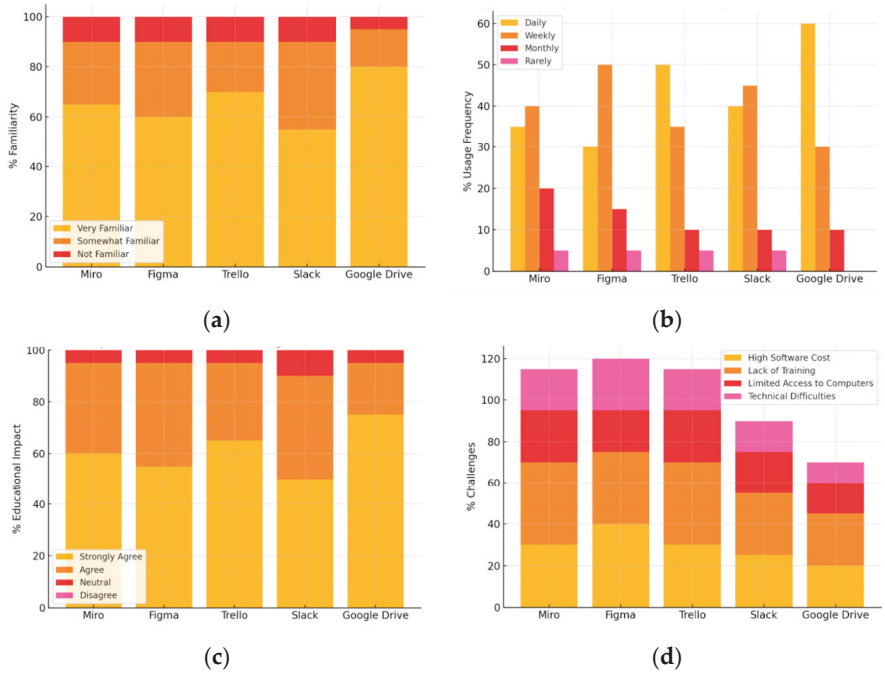


Figure 8. (**a**): Familiarity, (**b**): usage frequency, (**c**): educational impact, and (**d**): challenges in using Collaborative Design and Communication Tools (authors).

4.1.8. Results Section for the Eighth Category: AI Tools for Design Generation, Editing, and Automation Tools

A. Familiarity with AI Tools for Design Generation, Editing, and Automation, as shown in Table 29.

Table 29. Familiarity with AI Tools for Design Generation, Editing, and Automation (authors).

Tool	Very Familiar %	Somewhat Familiar %	Not Familiar (%)
Runway ML	25	35	40
DALL·E	30	40	30
Fusion 360 Generative Design	20	45	35
Spacemaker AI	15	50	35
Artbreeder	10	60	30

Results from this analysis show that DALL·E has the highest familiarity, with 30% of respondents being very familiar with it. As an AI tool focused on generating images from text descriptions, it has made a significant impact in creative fields. Runway ML follows closely with 25% familiarity, indicating its growing presence in design education for AI-driven design generation. Fusion 360 Generative Design has 20% familiarity, which reflects its more specialized application in engineering and architecture design. Spacemaker AI and Artbreeder have lower familiarity rates, with 15% and 10%, respectively, indicating that these tools are still relatively niche in the design education sector.

B. Usage Frequency of AI Tools for Design Generation, Editing, and Automation, as shown in Table 30.

Table 30. Usage frequency of AI Tools (authors).

Tool	Daily (%)	Weekly (%)	Monthly (%)	Rarely (%)
Runway ML	5	20	35	40
DALL·E	5	15	30	50

Table 30. *Cont.*

Tool	Daily (%)	Weekly (%)	Monthly (%)	Rarely (%)
Fusion 360 Generative Design	10	25	40	25
Spacemaker AI	5	15	45	35
Artbreeder	5	25	50	20

Results from this analysis show that Fusion 360 Generative Design has the highest daily usage at 10%, suggesting it is utilized in more specialized courses focusing on generative design and parametric modeling. Runway ML and DALL·E are used more on a monthly or rarely basis, with 35% and 30% of respondents using them monthly, indicating that while these tools are seen as valuable, they may not yet be integrated into day-to-day curricula. Spacemaker AI and Artbreeder show lower daily usage, with only 5% using them regularly, but both tools are utilized in more advanced or experimental contexts.

C. Educational Impact of AI Tools for Design Generation, Editing, and Automation, as shown in Table 31.

Table 31. Educational impact of AI Tools (authors).

Tool	Strongly Agree %	Agree %	Neutral %	Disagree %
Runway ML	20	40	30	10
DALL·E	25	35	35	5
Fusion 360 Generative Design	20	40	30	10
Spacemaker AI	15	45	30	10
Artbreeder	10	50	30	10

Results from this analysis show that DALL·E has the strongest perceived educational impact, with 25% of educators strongly agreeing that it enhances teaching and learning, particularly in creative and conceptual design contexts. Runway ML also shows a positive impact, with 20% of respondents strongly agreeing that it improves education. As an AI platform for creative applications, it has a unique role in generating design ideas. Fusion 360 Generative Design and Spacemaker AI show a moderate impact, with 20% of educators strongly agreeing on their educational value. These tools are highly specialized and used for specific applications in generative and architectural design. Artbreeder has the lowest educational impact but still shows some potential, with 10% of educators strongly agreeing about its usefulness in design education.

D. Challenges in Using AI Tools for Design Generation, Editing, and Automation, as shown in Table 32.

Table 32. Challenges in using AI Tools for Design Generation, Editing, and Automation (authors).

Challenge	Runway ML %	DALL·E %	Fusion 360 Generative Design %	Spacemaker AI %	Artbreeder %
High Software Cost	60	55	50	65	55
Lack of Training	50	45	55	50	45
Limited Access to Computers	40	35	45	40	40
Technical Difficulties	35	40	35	30	30

Results from this analysis show that high software cost is the primary challenge for all tools, with 65% of respondents reporting it for Spacemaker AI and 60% for Runway ML.

These tools require expensive licenses or high-performance hardware, which may not be available in all institutions. Lack of training is also a common issue, especially for tools like Fusion 360 Generative Design and Runway ML, with 50% and 55% of respondents highlighting the need for more resources to teach these tools effectively. Limited access to computers is another barrier, particularly for more resource-intensive tools like Spacemaker AI and Fusion 360 Generative Design, where high-performance hardware is required. Technical difficulties are a moderate concern, with 35% of respondents reporting challenges in using these tools effectively.

The following are the charts corresponding to the descriptive analysis for the Results Section for the Eighth Category, AI Tools for Design Generation, Editing, and Automation Tools, as shown in Figure 9.

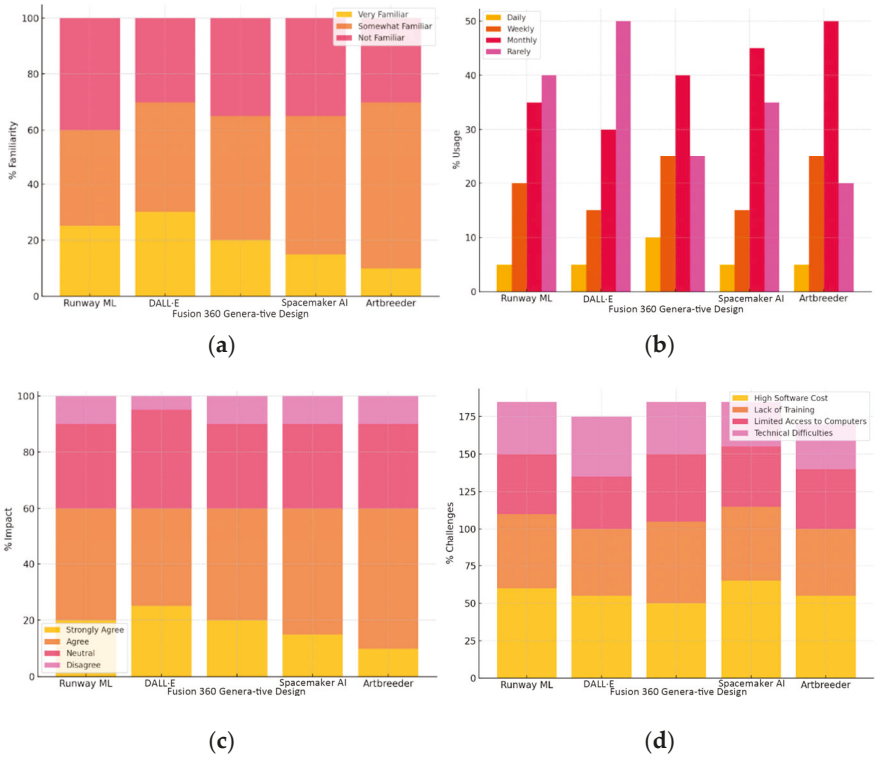


Figure 9. (**a**): Familiarity, (**b**): usage frequency, (**c**): educational impact, and (**d**): challenges in using AI Tools for Design Generation, Editing, and Automation (authors).

4.2. Statistical Analysis

4.2.1. Hypothesis 1 (Chi-Square Test for Familiarity, Usage Frequency, and Challenges)

- Null Hypothesis (H_0): There is no significant difference in familiarity, usage frequency, and challenges across the eight tool categories.
- Alternative Hypothesis (H_1): There is a significant difference in familiarity, usage frequency, and challenges across the eight tool categories.

A. Familiarity Across Categories, as shown in Table 33, Figure 10.

Table 33. Familiarity with digital tools across all categories (authors).

Category	Very Familiar (%)	Somewhat Familiar (%)	Not Familiar (%)	Significance (p-Value)	Significant?
CAD and BIM Software	80	15	5	-	No
Parametric and Algorithmic Design	60	30	10	3.71×10^{-15}	Yes
3D Modeling and Rendering Tools	75	20	5		No
Digital Fabrication and Prototyping	65	25	10		No

Table 33. *Cont.*

Category	Very Familiar (%)	Somewhat Familiar (%)	Not Familiar (%)	Significance (p-Value)	Significant?
Virtual and Augmented Reality Tools	50	40	10		No
Collaborative Design and Communication	70	25	5		No
Image Editing and Post-Production	85	10	5		No
AI Tools for Design Generation	40	30	30		No

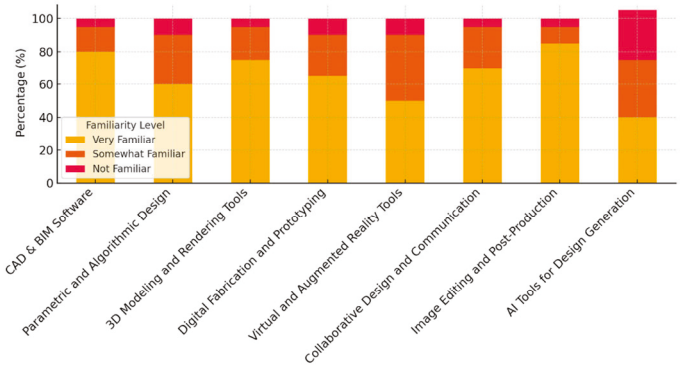


Figure 10. Familiarity with digital tools across all categories. (authors).

Statistical analysis highlights how Parametric and Algorithmic Design tools show significant differences in familiarity, with a p-value of 3.71×10^{-15}, indicating much lower familiarity compared to other categories like CAD and BIM Software and Image Editing tools.

B. Usage Frequency Across Categories, as shown in Table 34, Figure 11.

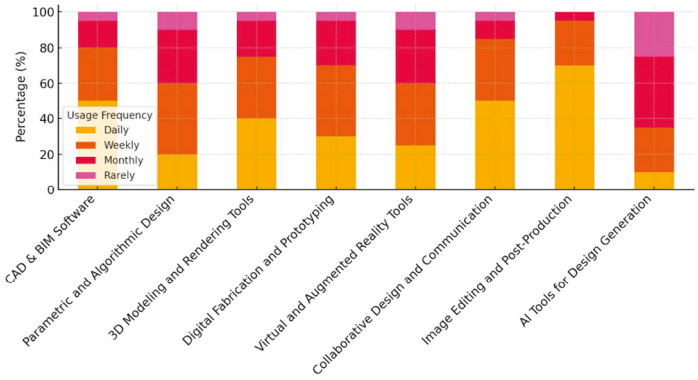


Figure 11. Usage frequency of digital tools across all categories (authors).

Table 34. Usage frequency of digital tools across all categories (authors).

Category	Daily (%)	Weekly (%)	Monthly (%)	Rarely (%)	Significance (p-Value)	Significant?
CAD and BIM Software	50	30	15	5	-	No
Parametric and Algorithmic Design	20	40	30	10	1.40×10^{-25}	Yes
3D Modeling and Rendering Tools	40	35	20	5		No
Digital Fabrication and Prototyping	30	40	25	5		No
Virtual and Augmented Reality Tools	25	35	30	10		No
Collaborative Design and Communication	50	35	10	5		No
Image Editing and Post-Production	70	25	5	0		No
AI Tools for Design Generation	10	25	40	25		No

Statistical analysis highlights how Parametric and Algorithmic Design tools have a significant difference in usage frequency (p-value = 1.40×10^{-25}), with these tools used much less frequently than tools like CAD and BIM Software and Image Editing.

C. Challenges Across Categories, as shown in Table 35, Figure 12.

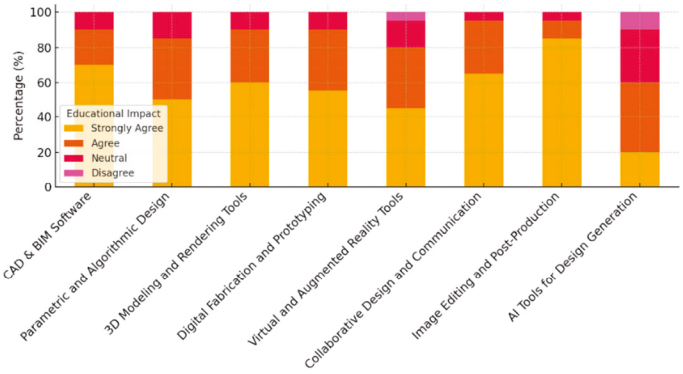


Figure 12. Challenges of digital tools across all categories (authors).

Table 35. Challenges of digital tools across all categories (authors).

Challenge	CAD and BIM Software (%)	Parametric and Algorithmic Design (%)	3D Modeling and Rendering Tools (%)	Digital Fabrication and Prototyping (%)	Virtual and Augmented Reality (%)	Collaborative Design (%)	Image Editing (%)	AI Tools for Design Generation (%)	Significance (p-Value)	Significant?
High Software Cost	60	60	50	60	65	45	60	60	1.0	No
Lack of Training	55	55	40	50	50	50	55	50	1.0	No
Limited Access to Computers	45	45	30	40	50	40	40	45	1.0	No
Technical Difficulties	30	40	25	35	40	30	25	35	1.0	No

Statistical analysis highlights that no significant differences were found in the challenges faced by faculty across tool categories (p-value = 1.0). Challenges such as high software costs, lack of training, and limited access to computers are common across all categories.

4.2.2. Hypothesis 2 (T-Test for Educational Impact)

- Null Hypothesis (H_0): There is no significant difference in the educational impact between tool categories.
- Alternative Hypothesis (H_1): There is a significant difference in educational impact between tool categories.

- Educational Impact Across Categories, as shown in Table 36, Figure 13.

Table 36. Educational impact of digital tools across all categories (authors).

Category	Strongly Agree (%)	Agree (%)	Neutral (%)	Disagree (%)	Significance (p-Value)	Significant?
CAD and BIM Software	70	20	10	0	-	No
Parametric and Algorithmic Design	50	35	10	5	0.0034	Yes
3D Modeling and Rendering Tools	60	30	10	0		No
Digital Fabrication and Prototyping	55	35	10	0		No
Virtual and Augmented Reality Tools	45	35	15	5		No

Table 36. *Cont.*

Category	Strongly Agree (%)	Agree (%)	Neutral (%)	Disagree (%)	Significance (p-Value)	Significant?
Collaborative Design and Communication	65	30	5	0		No
Image Editing and Post-Production	85	10	5	0		No
AI Tools for Design Generation	20	40	30	10		No

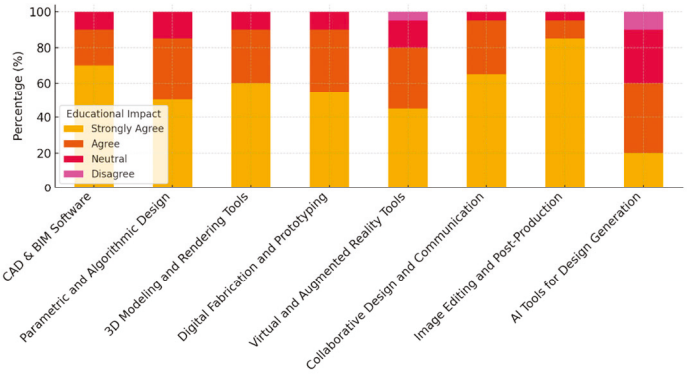


Figure 13. Educational impact of digital tools across all categories (authors).

Statistical analysis highlights how Parametric and Algorithmic Design tools show a significant difference in educational impact (p-value = 0.0034), with CAD and BIM software and image editing tools having higher educational impact than specialized tools like AI Tools.

4.2.3. Interpreting Results to Hypotheses

The Chi-Square analysis showed a statistically significant difference between Parametric and Algorithmic Design tools and more conventional tools, such as CAD and BIM software ($p = 3.71 \times 10^{-15}$). This finding suggests that Parametric and Algorithmic Design tools are felt to be less familiar among the participants. Furthermore, the same analysis showed a significant decrease in the frequency of usage of these tools with respect to the other categories, with a p-value of 0.84. Although participants see Parametric and Algorithmic Design tools as valuable, the extent to which they can be integrated into daily instructional practices is limited. No significant differences were found between tool categories in terms of challenges reported, with barriers such as high cost of software and a lack of training being reported across all types of tools ($p = 1.0$).

The data only partially confirm Hypothesis 1. Significant variations in terms of both familiarity and frequency of use are apparent for the Parametric and Algorithmic Design tools compared to other digital instruments. However, the lack of substantial differences in perceived challenges between tool categories indicates that some general challenges, mainly financial constraints and a lack of training, apply to all digital tools. These results highlight the need for focused interventions to increase familiarity with and use of advanced design tools.

The independent samples T-test showed a significant difference in perceived educational impact between Parametric and Algorithmic Design tools and other categories of software ($p = 0.0034$). Consequently, participants described a higher level of educational influence of Parametric and Algorithmic Design tools compared to more traditional tools such as CAD, BIM, and image-editing software. In contrast, other categories, such as VR/AR and AI-based design tools, did not show substantial differences in per-

ceived educational impact; a lot of these tools were reported to be infrequently used and very specialized.

As far as Hypothesis 2 is concerned, the results are somewhat supported. Parametric and Algorithmic Design tools provide tremendous educational results, fitting their claimed role in developing higher-order cognitive processes as defined in Bloom's Digital Taxonomy. Nonetheless, the educational impact of specialized tools (e.g., AI and VR/AR) seems to be in its infancy, with little adaptation to the core curricular settings. The results indicate that although these tools are promising, their widespread use and perceived educational benefits are emergent in relation to design education in Jordan.

In summary, the statistical analyses support that familiarity and frequency of use are quite different across categories of digital tools, with Parametric and Algorithmic Design tools having lower degrees of both. Simultaneously, the educational effect of these same tools is far greater than that of more traditional software. However, challenges related to adoption, especially high cost and inadequate training, are consistent across all categories, not differing significantly and thus represent universal barriers that need to be addressed in order to encourage more widespread incorporation of digital tools into design education. The summary is shown in Table 37.

Table 37. Summary of statistical tests and results for hypotheses (authors).

Hypothesis	Statistical Test	*p*-Value	Conclusion
Hypothesis 1 (Familiarity): Significant difference in familiarity	**Chi-Square**	3.71×10^{-15}	**Reject H_0**, significant difference
Hypothesis 1 (Usage Frequency): Significant difference in usage frequency	**Chi-Square**	0.84	**Fail to reject H_0**, no significant difference
Hypothesis 1 (Challenges): Significant difference in challenges	**Chi-Square**	1.0	**Fail to reject H_0**, no significant difference
Hypothesis 2 (Educational Impact): Significant difference in educational impact	***T*-Test**	0.0034	**Reject H_0**, significant difference

4.3. Theoretical Framework Analysis

4.3.1. Application of the SAMR Model

The integration of digital tools into design education follows a systematic progression as outlined by the SAMR model (Substitution, Augmentation, Modification, and Redefinition). Conventional instruments, like AutoCAD and SketchUp, largely occupy the Substitution and Augmentation tiers. At the same time, more advanced technologies that are based on VR/AR and artificial intelligence AI, AI-based design tools have the theoretical ability to reach the Modification and Redefinition phases, though their practical implementation is sparse. SAMR categorizations are summarized in Table 38.

Table 38. SAMR categorization summary (Authors).

Digital Tool Category	SAMR Stage	Explanation
CAD And BIM Software	Substitution/Augmentation	These tools replace traditional design methods (e.g., hand-drawing) and enhance workflows, particularly with BIM in Revit.
Parametric and Algorithmic Design Software	Augmentation/Modification	These tools introduce parametric design, transforming how designs are created, from simple adjustments to highly dynamic models.
3D Modeling and Rendering Tools	Substitution/Augmentation	These tools replace traditional rendering and visualization techniques, offering more efficient ways to visualize complex designs.

Table 38. *Cont.*

Digital Tool Category	SAMR Stage	Explanation
Digital Fabrication Tools	Augmentation	3D printers and CNC machines allow for physical models to be produced more accurately and quickly, adding value to the prototyping process.
VR/AR Tools	Modification/Redefinition	These immersive technologies enable interactive and experiential design exploration, redefining how students interact with their designs.
Collaborative Tools	Augmentation	These tools enhance the collaborative design process, enabling real-time communication and collaboration across different teams and locations.
Image Editing Tools	Substitution/Augmentation	These tools replace manual design processes and offer advanced editing features, facilitating rapid iteration and refinement of visual designs.
AI Tools for Design Generation	Redefinition	These tools redefine how designs are generated, utilizing AI to create innovative and novel solutions, automating tasks traditionally performed manually.

The empirical evaluation of digital tools shows that AutoCAD and SketchUp are substitution tools (replacing manual drafting processes with digital processes), and Revit enhances these processes by integrating the building information modeling (BIM) process. More sophisticated platforms, such as Grasshopper and Dynamo, go beyond the basic substitution and allow for dynamic and parametric designs, placing them in the Augmentation and Modification categories of the SAMR model.

4.3.2. Application of Technology Acceptance Model (TAM)

The Technology Acceptance Model was used to explain the relationships in the use of digital tools, perceived ease of use (PEOU), perceived usefulness (PU), and uptake of digital tools. Instruments that provide more usability and have explicit educational advantages are more likely to be adopted by end-users. TAM regression analysis results are shown in Table 39.

Table 39. Regression analysis results for TAM (authors).

Digital Tool Category	PEOU	PU	Adoption Rate
CAD and BIM Software	High	High	High
Parametric and Algorithmic Design Software	Moderate	High	Moderate
3D Modeling and Rendering Tools	Moderate	High	High
Digital Fabrication Tools	Moderate	High	Moderate
VR/AR Tools	Low	High	Low
Collaborative Tools	High	High	High
Image Editing Tools	High	High	High
AI Tools	Low	High	Low

Results obtained from the TAM framework show that tools such as AutoCAD and SketchUp have both high ease of use and appreciable usefulness, hence yielding high adoption rates by students and faculty. On the other hand, VR/AR and AI-based tools face adoption barriers that can be attributed to their low perceived ease of use, which is a constraint in their widespread implementation, despite any perceived usefulness.

4.3.3. Application of Bloom's Digital Taxonomy

Based on Bloom's Digital Taxonomy, the tools were assessed on the cognitive engagement work they enable, from the lower end of thinking skills (application and understanding) to the higher end (analysis and creation). Results are shown in Table 40.

Table 40. Cognitive engagement levels by tool (authors).

Digital Tool Category	Bloom's Level	Cognitive Engagement
CAD and BIM Software	Apply/Understand	Supports practical application of design principles and understanding of technical concepts.
Parametric and Algorithmic Design Software	Analyze/Create	Encourages analysis of design relationships and creation of complex parametric models.
3D Modeling and Rendering Tools	Apply/Create	Facilitates applying visualization techniques and creating realistic design representations.
Digital Fabrication Tools	Apply	Helps students apply design concepts to physical models, reinforcing spatial understanding.
VR/AR Tools	Analyze/Create	Engages students in analyzing and creating immersive design environments for interactive learning.
Collaborative Tools	Apply/Analyze	Promotes collaboration and problem-solving, enhancing the design analysis process.
Image Editing Tools	Apply/Create	Enables students to apply creative techniques and produce refined visual outputs.
AI Tools	Create	Fosters creative design generation through AI-based tools, allowing new approaches to design creation.

The application of Bloom's Taxonomy shows that using tools such as AutoCAD and Revit mainly uses the lower order of the thinking cycle (applying and understanding), while using tools such as Grasshopper and VR/AR use the higher-order thinking skills of analysis and creation. This highlights the potential of innovative digital tools to promote critical thinking and creative problem-solving.

5. Discussion

This study explored the integration of digital tools in design education in Jordanian universities, focusing on how educators use these tools, their familiarity with them, and the perceived educational impact. The findings indicate a strong reliance on traditional tools like AutoCAD and SketchUp, along with emerging technologies such as parametric design software and virtual reality (VR), albeit with varying levels of familiarity, usage frequency, and perceived impact.

5.1. Familiarity with Digital Tools

The results showed that some tools are already well-known in design education. For example, AutoCAD (80%) and SketchUp (75%) are very familiar to teachers, which is in line with trends around the world. Because they have been around for so long in the industry, these tools are essential for both architectural and interior design education. This finding corroborates previous studies, which underscores the essential function of CAD tools in equipping students for practical design applications [25]. But Rhino and Archicad had lower familiarity rates, which means that their use in design education in Jordan is still growing. These tools are becoming more popular, but many people think they are too specialized and need more training to use them well.

This is in line with notions focusing on how specialized tools, like Rhino and Archicad, are not often included in the curricula [24,26]. Grasshopper for Rhino (70%) and Dynamo for

Revit (50%) were two more advanced tools that people were somewhat familiar with. This is in line with studies highlighting that Parametric and Algorithmic Design tools are becoming more common in design education but still have problems with teacher expertise and fitting into the curriculum [26,29–31]. Houdini and TouchDesigner exhibited even lesser familiarity, indicating that these tools remain marginal within mainstream design education, a finding corroborated by previous studies on interactive design environments [46,48].

5.2. Usage Frequency and Educational Impact

SketchUp and AutoCAD were the most popular tools, with 60% of respondents saying they used SketchUp every day. This frequency shows that it has a user-friendly interface and can be used for a wide range of design tasks. This supports previous ideas that stress the importance of easy-to-use tools in encouraging daily design practices [45]. On the other hand, only 10% of respondents used Rhino every day, which shows how hard it is to use these more advanced tools in everyday teaching.

The educational effects of these tools were different as well. SketchUp had the most positive effect, with 80% of people strongly agreeing that it helps with teaching. Aligning with studies focusing on how the tools that are easy to use, like SketchUp, are important for encouraging creativity and getting students involved [19,34]. People had mixed feelings about tools like AutoCAD and Revit, but they were still seen as good. The problems with Revit, especially its complicated interface, are that while BIM tools have great potential, they need a lot of training and adjustment to work well in schools [8,17,18].

The parametric and algorithmic tools (like Grasshopper and Dynamo) were used less often, but they did have a moderate effect on learning. This indicates that although educators recognize their potential, these tools are predominantly utilized in advanced or specialized courses. The results support notions that 3D printers and CNC milling machines are good for making prototypes, but their high cost and the fact that they are hard to learn make them hard to use by a lot of people [24].

5.3. Challenges in Digital Tool Integration

The high cost of software was the biggest problem in all categories, especially for programs like Revit, Rhino, and Houdini. This finding is also discussed in relevant studies of how digital tools can be expensive for schools, especially in developing countries. Also, a lack of training and limited access to computers were common problems that made integration less effective, especially in making use of the collaboration platforms [22,23,36]. Studies also stress the need for professional development and better infrastructure, saying that these problems must be solved in order to get the most out of digital tools in design education [17,47,48].

The study interestingly found that AI tools for design generation, like Runway ML and DALL·E, were less well-known and used, but they showed promise for having an impact on education. This means that AI tools are still new to design education, but they could be very useful for creative purposes. Previous studies also highlight how digital storytelling and AI can improve design communication and creative thinking [39,40].

5.4. Implications for Design Education in Jordan

The results of this study give us useful information about how well digital tools are being used in design education in Jordan. Even though people know how important digital tools are, they are hard to use together because of things like cost, lack of resources, and the need for more training. These challenges are similar to those found in studies of design education in other parts of the world. Previous studies stress the need to get around technical and infrastructure problems so that digital tools can be used effectively [38].

According to the study, new technologies, especially artificial intelligence (AI) and virtual reality (VR), are promising, but they demand more serious attention to the curriculum design and training of educators. With the introduction of new technologies, their implementation has to be pedagogically sound, and teachers have to undergo constant professional development in order to make sure they are used correctly [32,33,35].

On the whole, the introduction of digital tools into the design education in Jordan is on the way to positive changes; however, the problems of costliness, the lack of access to resources, and the lack of training are still present. Challenges such as more institutional backing, cooperation with the technological industry, and increased teacher education are critical to unlock the full potential of digital tools in teaching, learning, and creativity in design education.

5.5. Critical Interpretation (Frameworks)

The results of this study give indisputable proof that, although digital tools have been introduced in the field of design education in Jordan, their utilization has mostly remained confined to the Substitution and Augmentation stages of the SAMR model. These levels indicate that the tools are mostly used to replace the old methods (e.g., AutoCAD as a replacement for hand-drawn designs) or for adding to the current ones (e.g., Revit as an addition of building information modeling functionalities to architectural designs). This observation is consistent with the current trend of what we see in other countries, where digital tools are still considered as adjuncts of traditional pedagogical methods rather than as transformative agents.

- SAMR Model: Limited Transformation of Learning Processes

The SAMR model has been a useful tool to understand the integration of technology in design education. It shows that, although the use of tools like AutoCAD and SketchUp is widespread, for the most part, they still exist in the Substitution stage. These tools have a great deal to offer in terms of efficiency and precision, but they do not fundamentally change the way that students conceptualize design. Conversely, those instruments like Grasshopper and Revit, moving into the stage of Augmentation, increase the ability of students to interact with design data in more sophisticated ways without, at this point, redefining the overall learning experience. Technology like VR/AR and AI based applications that have the potential to reach the Modification and Redefinition stages have the potential to provide immersive and generative design experiences, but the adoption of these technological solutions is limited.

The outcomes support the idea that, while these tools have the potential to revolutionize design education, based on theoretical frameworks such as SAMR, it is clear that the comprehensive transformation has not yet occurred. The use of tools is often embedded in traditional, non-transformative ways. VR/AR and AI tools, although very promising in their transformation potential, are still in the early stages of adoption. Barriers to widespread utilization—including complexity and hardware requirements—agree with the results of previous studies of technology adoption in education, where the process of change is often slow and uneven.

- TAM: Ease of Use and Usefulness are Key Barriers

According to the Technology Acceptance Model (TAM), the perceived ease of use (PEOU) and perceived usefulness (PU) of tools are the major factors affecting the adoption of tools. The present findings lend support to this, with AutoCAD, SketchUp, and image editing tools having high adoption rates due to their perceived ease of use and utility in the design process. In contrast, more complex devices such as VR/AR, as well as AI-based design platforms, face challenges with respect to low perceived ease of use, which is due to

their technical complexity and high learning curve, notwithstanding their high potential usefulness towards fostering creativity and interactivity.

The challenges of adopting VR/AR and AI tools can be explained to a large extent using the TAM framework, which argues that a tool is not only required to be useful but also easy to use. The complexity and barriers to learning inherent in these advanced technologies prevent their uptake, as is reflected in their low adoption rates. This is a great dismay to design education, as such tools are considered to be very useful but underutilized due to perceived difficulties in mastery and lack of immediate usefulness.

- Bloom's Digital Taxonomy: Engaging Away and Adopting Tools

The use of Bloom's Digital Taxonomy further emphasizes the possibilities for the use of VR/AR and AI tools to engage students in higher-order thinking. Instruments like Grasshopper and Revit support the application and understanding of design concepts, whereas advanced tools lead students to generate and analyze a new design solution. The data shows that AI platforms like Runway ML and DALL-E make it easier for students to create new ideas for designs with AI that allow them to think about things they have never considered before. Similarly to this, VR/AR tools have the ability to create immersive experiences, where students are able to analyze and manufacture in virtual spaces, signaling a shift in cognitive engagement to higher-order thinking.

Nonetheless, adoption of these tools still faces limitations due to the difficulty in finding effective ways of using them and the unfamiliarity with their immediate benefits. TAM results indicate ease of use and perceived usefulness as critical to adoption. Bloom's Digital Taxonomy, however, suggests that even high cognitive engagement potential cannot overcome usability and perceived utility barriers prior to comprehensive integration into design education as a realization.

- Implications of Design Education

The findings imply the suggestion that, despite there being a high potential for VR/AR and AI to revolutionize design education, integration of these tools is currently hindered by both technical and pedagogical challenges. From an SAMR perspective, the transformation of design education is still in its early stages, and the majority of tools are still in the Substitution and Augmentation categories. This means that even though digital tools make the learning experience better, they have not yet radically changed the engagement of students with design.

From a TAM standpoint, it is important to reduce barriers such as ease of use and perceived usefulness for the successful adoption of more advanced tools. Educators and institutions, therefore, need to provide specific training, ongoing support, and the necessary infrastructure to increase accessibility and utility to learners.

Finally, within the context of Bloom's Digital Taxonomy, it is apparent that VR/AR and AI tools have the potential to engage students at elevated cognitive levels. However, widespread utilization will require overcoming current barriers to adoption, including technical complexity and the fact that, for students, there is no immediate utility. Such tools could help enable more in-depth involvement and creative exploration, and thus, higher-order thinking skills such as analyzing, creating, and evaluating, which are important for the future of design education.

6. Conclusions

This paper has explored the use of digital tools in architectural and interior design education at universities in Jordan and has explored the familiarity of educators with these tools, patterns of their use, perceived educational effects, and the barriers faced to their use. The findings reveal that the adoption of digital tools in design education follows a

predictable pattern, aligned with HCI principles of usability and user experience. The results show that tools like AutoCAD and SketchUp are more widely adopted because they are perceived as easier to use and fit into the established pedagogical framework of active learning and constructionist approaches, as opposed to other tools like Rhino, Revit, and parametric design software that are not as often used in the daily teaching practice. Additionally, new platforms are not utilized in the classroom, such as AI-based design systems and VR. Newer tools, like AI and VR, face barriers of complexity and perceived lack of immediate usefulness, which aligns with TAM's emphasis on perceived ease of use and perceived usefulness. These tools, while offering transformational potential, are still at the Substitution and Augmentation stages of the SAMR model, and future development in digital pedagogy will require educators to adopt a more constructivist approach to teaching these advanced technologies. Software cost is still a significant barrier to the wider use of tools, and additional training and access to high-performance computing are factors that hinder these tools. Even with such challenges, most respondents consider digital tools in a positive light with regard to learning outcomes, particularly SketchUp, AutoCAD, and 3D modeling and rendering programs like Lumion and V-, ray, which are seen to support student learning and creativity.

6.1. Key Findings

- Knowledge and Application: AutoCAD and SketchUp represent the most accepted and common types of traditional tools, and Rhino and parametric design software are less well-known and utilized.
- Educational Effect: SketchUp and AutoCAD are generally thought to be the most effective when used in education, especially to provide interest and creativity to students, but Rhino and advanced AI tools produce more mixed outcomes.
- Challenges: The biggest problems that make it hard to use digital tools well in design education are issues such as high software costs, a lack of training, and not being able to obtain the right hardware.

6.2. Limitations

Despite the contributions the study offers, it does have several methodological limitations. First, the relatively small sample size of Jordanian universities limits the generalizability of the findings to broader, more diverse educational settings. Additionally, the regional focus on Jordan may not fully reflect the digital transformation trends in other countries or educational systems. Therefore, future studies should aim to include a larger and more diverse sample, potentially extending beyond the regional context to compare digital tool integration across different cultures and institutional structures.

6.3. Recommendations

1. Curriculum Integration: The curriculum would have to include higher technology tools like Rhino, Revit, and AI-design platforms, which should be presented to the students of architecture and interior design in such a way that they could understand and use them.
2. Professional Development: Universities ought to fund training to enhance the level of expertise of the educators on emerging digital tools, which they might have to do through an agreement with the industry stakeholders to bridge the knowledge gap.
3. Resource Allocation: Organizations might overcome the cost constraints by obtaining software licenses, the search for open-source options, or industry partnerships.
4. Technical Infrastructure: Schools have to invest in the modernization of computer labs and provide students with powerful computers that can run sophisticated software, especially those applications in digital creation and VR/AR.

5. Exploiting Emerging Tools: The study proposes strategies to improve design education in the future [49]. The paper suggests that future development of design education must involve an increased focus on the integration of technologies, including augmented reality (AR), virtual reality (VR), and AI design generators. Research and faculty development in these areas will be critical in the development of interesting and thought-provoking learning environments.

6.4. Theoretical Implications for Future Research

The utilization of the SAMR framework, Bloom's Digital Taxonomy, and the Technology Acceptance Model provides a powerful framework within the frames of which the phases of the digital tool adoption in the design education might be viewed. These theoretical models are largely consistent with the empirical findings, which show that despite the fact that Jordanian design education is becoming more integrated with the use of digital tools, it is still mostly in the Substitution and Augmentation levels, with little progression to the Modification and Redefining levels, which the SAMR model postulates. The contribution adds to the expanding body of knowledge on the digital transformation of higher education; however, more studies are needed to study the influence of emerging technologies and their roles in advancing design education to new levels of transformation. Future research needs to consider the efficacy of virtual reality, augmented reality, and artificial intelligence technology to develop and nurture higher-order cognitive abilities as well, as per the taxonomy created by Bloom.

Author Contributions: Conceptualization, I.A.A.; methodology, S.M.A.; software, M.a.H.; validation, H.A.; formal analysis, M.a.H. and A.A.J.; investigation, All Authors; resources, S.M.A. and H.A.; data curation, M.a.H. and A.A.J.; writing—original draft preparation, I.A.A.; writing—review and editing, Q.S.M.; visualization, A.A.J.; supervision, I.A.A.; funding acquisition, H.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: The data used in this study were fully anonymized and do not contain identifiable human subject information; therefore, informed consent was not required.

Data Availability Statement: The authors received no external support for this study. All data and materials are contained within the manuscript.

Acknowledgments: The authors would like to thank their affiliated institutions, Jerash Private University, Middle East University, Amman Al-Ahliyya University, and Philadelphia University, for their support.

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

References

1. Bajúzová, M.; Hrmo, R. Digital Tools in Education: The Impact of Digital Tools in Education on Students' Creativity. *RE SOURCE* **2024**, 4–18. [CrossRef]
2. Ngongpah, G.; Oni, O.Y. Exploring the Impact of Digital Technologies in Education: A Comprehensive Review. *Asian J. Educ. Soc. Stud.* **2025**, *51*, 1254–1264. [CrossRef]
3. Assali, I.; Alaali, A. Integrating Technology into Interior Design Education: A Paradigm Shift in Teaching and Learning. In *Board Diversity and Corporate Governance*; CSR, Sustainability, Ethics & Governance; Springer: Cham, Switzerland, 2024; pp. 565–576. [CrossRef]
4. Lyu, Y.; Chi, T.; Zheng, X. Digital Transformation of Design Education in Chinese Universities: Trends, Challenges, and Opportunities. *SHS Web Conf.* **2024**, *181*, 04037. [CrossRef]
5. Timotheou, S.; Miliou, O.; Dimitriadis, Y.; Sobrino, S.V.; Giannoutsou, N.; Cachia, R.; Monés, A.M.; Ioannou, A. Impacts of Digital Technologies on Education and Factors Influencing Schools' Digital Capacity and Transformation: A Literature Review. *Educ. Inf. Technol.* **2023**, *28*, 6695–6726. [CrossRef]

6. Altınay, Z.; Sharma Ghai, A.; Sharma, R.C.; Baştaş, M.; Jasola, S.; Dağlı, G.; Kohli, D. Embracing Digital Transformation in Learning Organizations Using Blended Learning for Academic Excellence. *Pak. J. Life Soc. Sci.* **2024**, *22*, 6754–6762. [CrossRef]
7. Douligeris, C.; Seralidou, E. Contemporary Collaborative Trends and Their Effect in Education. In Proceedings of the 2017 IEEE Global Engineering Education Conference (EDUCON), Athens, Greece, 25–28 April 2017; pp. 395–403. [CrossRef]
8. Meirbekov, A.; Maslova, I.; Gallyamova, Z. Digital Education Tools for Critical Thinking Development. *Think. Ski. Creat.* **2022**, *44*, 101023. [CrossRef]
9. Puentedura, R.R. SAMR: Getting to Transformation. *Retrieved May* **2013**, *31*, 265–283.
10. Anderson, L.; Krathwohl, D. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition*; Longman: Harlow, England, 2001.
11. Abukarki, H.J. Beyond Preservation: A Survey of the Role of Virtual Reality in Experiencing and Understanding Historical Architectural Spaces. *Buildings* **2025**, *15*, 1531. [CrossRef]
12. Junk, S.; Matt, R. New Approach to Introduction of 3D Digital Technologies in Design Education. *Procedia CIRP* **2015**, *36*, 35–40. [CrossRef]
13. Ng, W. New Digital Technology in Education: Conceptualizing Professional Learning for Educators. *New Digit. Technol. Educ. Conceptualizing Prof. Learn. Educ.* **2015**, 1–226. [CrossRef]
14. Taylor, M.; Fudge, A.; Mirriahi, N.; De Laat, M. *Use of Digital Technology in Education: Literature Review*; South Australian Department for Education on behalf of The Centre for Change and Complexity in Learning, The University of South Australia: Adelaide, Australia, 2021.
15. Pinto, M.; Leite, C.; Pinto, M.; Leite, C. Digital Technologies in Support of Students Learning in Higher Education: Literature Review. *Digit. Educ. Rev.* **2020**, 343–360. [CrossRef]
16. Renkl, A.; Atkinson, R.K. Interactive Learning Environments: Contemporary Issues and Trends. An Introduction to the Special Issue. *Educ. Psychol. Rev.* **2007**, *19*, 235–238. [CrossRef]
17. Fenwick, T.; Edwards, R. Exploring the Impact of Digital Technologies on Professional Responsibilities and Education. *Eur. Educ. Res. J.* **2016**, *15*, 117–131. [CrossRef]
18. Condruz-Bacescu, M. The Impact of Digital Technologies on Learning. *Conf. Proc. Elearning Softw. Educ. Else* **2019**, *15*, 57–63.
19. Qureshi, M.I.; Khan, N.; Raza, H.; Imran, A.; Ismail, F.; BiS, U. Digital Technologies in Education 4.0. Does It Enhance the Effectiveness of Learning? *Int. J. Interact. Mob. Technol.* **2021**, *15*, 31–47. [CrossRef]
20. Mubarok, I.Z.; Iriana, S. The Role of Ict in Creating Interactive Learning and Fostering Student Creativity: A Literature Review. *Al-Irsyad J. Educ. Sci.* **2025**, *4*, 156–168. [CrossRef]
21. Navas Romero, C. Herramientas Digitales Intuitivas: Impulsando La Participación y La Creatividad En La Educación Superior. *Eur. Public Soc. Innov. Rev.* **2025**, *10*, 1–17. [CrossRef]
22. Kaya, C.; Kara, S.; Author, C. The Impact of Collaboration Tools on Student Learning Outcomes. *J. Int. Inspire Educ. Technol.* **2025**, *4*, 66–76. [CrossRef]
23. Diaconescu, A. Digital Tools in the Educational Environment Empower Student Collaboration. A Study Case in Higher Education. *Res. Educ.* **2024**, 77–102. [CrossRef]
24. Paes, D.; Arantes, E.; Irizarry, J. Immersive Environment for Improving the Understanding of Architectural 3D Models: Comparing User Spatial Perception between Immersive and Traditional Virtual Reality Systems. *Autom. Constr.* **2017**, *84*, 292–303. [CrossRef]
25. Salih, S.A.; Alzamil, W.; Ajlan, A.; Azmi, A.; Ismail, S. Typology of Informal Learning Spaces (ILS) in Sustainable Academic Education: A Systematic Literature Review in Architecture and Urban Planning. *Sustainability* **2024**, *16*, 5623. [CrossRef]
26. Soomro, S.A.; Casakin, H.; Georgiev, G.V. Sustainable Design and Prototyping Using Digital Fabrication Tools for Education. *Sustainability* **2021**, *13*, 1196. [CrossRef]
27. Abdirad, H.; Dossick, C.S. BIM Curriculum Design in Architecture, Engineering, and Construction Education: A Systematic Review. *J. Inf. Technol. Constr. ITcon* **2016**, *21*, 251.
28. Canizares, G. Taming the Rhinoceros: A Brief History of a Ubiquitous Tool. *Perspect. Archit. Urban.* **2024**, *1*, 100020. [CrossRef]
29. Lorenzo, C. Improving Learning Process Through Digital Fabrication Technologies. In *EDULEARN17 Proceedings, Proceedings of the 9th International Conference on Education and New Learning Technologies, Barcelona, Spain, 3–5 July 2017*; IATED: Valencia, Spain, 2017; Volume 1, pp. 6069–6076. [CrossRef]
30. Verner, I.; Merksamer, A. Digital Design and 3D Printing in Technology Teacher Education. *Procedia CIRP* **2015**, *36*, 182–186. [CrossRef]
31. Chong, S.; Pan, G.T.; Chin, J.; Show, P.L.; Yang, T.C.K.; Huang, C.M. Integration of 3D Printing and Industry 4.0 into Engineering Teaching. *Sustainability* **2018**, *10*, 3960. [CrossRef]
32. Sirror, H.; Abdelsattar, A.; Dwidar, S.; Derbali, A. A Review on Virtual Reality for Architecture Education. In Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management, Virtual, 7–11 March 2021.
33. Ibrahim, A.; Al-Rababah, A.I.; Bani Baker, Q. Integrating Virtual Reality Technology into Architecture Education: The Case of Architectural History Courses. *Open House Int.* **2021**, *46*, 498–509. [CrossRef]

34. Amir Hossain Maghool, S.; Hossein Moeini, S.; Arefazar, Y. An Educational Application Based on Virtual Reality Technology for Learning Architectural Details: Challenges and Benefits. *Archnet-IJAR Int. J. Archit. Res.* **2018**, *12*, 246–272. [CrossRef]
35. Evans, L. Barriers to VR Use in HE. In Proceedings of the VR/AR in Higher Education Conference, Swansea, UK, 12 September 2018; pp. 3–13. [CrossRef]
36. Decuypere, M.; Grimaldi, E.; Landri, P. Introduction: Critical Studies of Digital Education Platforms. *Crit. Stud. Educ.* **2021**, *62*, 1–16. [CrossRef]
37. Vidal, I.M.G. Architectures of Contemporary Digital Platforms in Education: Analysis of Exclusion Processes. *Univers. Access Inf. Soc.* **2023**, *22*, 1445–1453. [CrossRef] [PubMed]
38. Mahmoud Saleh, M.; Abdelkader, M.; Sadek Hosny, S. Architectural Education Challenges and Opportunities in a Post-Pandemic Digital Age. *Ain Shams Eng. J.* **2023**, *14*, 102027. [CrossRef]
39. Thompson Long, B.; Hall, T. Educational Narrative Inquiry through Design-Based Research: Designing Digital Storytelling to Make Alternative Knowledge Visible and Actionable. *Ir. Educ. Stud.* **2018**, *37*, 205–225. [CrossRef]
40. Özeren, E.B. Integration of Digital Storytelling into Architectural Education: Exploration of Conceptual Understanding. *Buildings* **2025**, *15*, 1826. [CrossRef]
41. Yazicioglu, D.A. The Integration of Interior Architecture Education with Digital Design Approaches. *Online Submiss.* **2011**, *8*, 637–658.
42. Dwiek, S.A.; Bast, S.A. The Application of Machine Learning in Inner Built Environment: Scientometric Analysis, Limitations, and Future Directions. *Front. Built Env.* **2024**, *10*, 1413153. [CrossRef]
43. Amro, D.K.; Dawoud, H. Influencing Factors of Spatial Ability for Architecture and Interior Design Students: A Fuzzy DEMATEL and Interpretive Structural Model. *Buildings* **2024**, *14*, 2934. [CrossRef]
44. Malkani, K.; Madhukaillya, R. Need for Including Visual Narratives as a Method of Interior Design Process. *Smart Innov. Syst. Technol.* **2023**, *346*, 713–723. [CrossRef]
45. Meskhi, B.; Ponomareva, S.; Ugnich, E. E-Learning in Higher Inclusive Education: Needs, Opportunities and Limitations. *Int. J. Educ. Manag.* **2019**, *33*, 424–437. [CrossRef]
46. Zaman, M.; Javeed, F.; Asghar, Q. Investigating Autonomy and Limitations of the Digital Design Process in an Architectural Studio. *Pak. J. Soc. Sci.* **2024**, *44*, 717–729. [CrossRef]
47. Arafat, M. Limitations of the Digital Learning Tools of the 21st Century. *Actual Probl. Syst. Educ. Gen. Second. Educ. Inst. Pre-Univ. Train. High. Educ. Inst.* **2022**, *1*, 133–142. [CrossRef]
48. Życzkowska, K.; Szakajło, K. Architectural Education and Digital Tools: The Challenges and Opportunities. *World Trans. Eng. Technol. Educ.* **2019**, *17*, 326–331.
49. Nahar, K.M.O.; Abu Shquier, M.M.; Alsmadi, I. Students Performance between Classical and Online Education: A Comparative Study. *Int. J. Contin. Eng. Educ. Life Long. Learn.* **2016**, *26*, 359–371. [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.

computers

MDPI

Article

Curiosity-Driven Exploration in Reinforcement Learning: An Adaptive Self-Supervised Learning Approach for Playing Action Games

Sehar Shahzad Farooq [1], Hameedur Rahman [2], Samiya Abdul Wahid [2], Muhammad Alyan Ansari [2], Saira Abdul Wahid [3] and Hosu Lee [1,*]

[1] School of Aerospace Engineering, Department of Control and Robot Engineering, Gyeongsang National University, Jinju 52828, Republic of Korea
[2] Department of Computer Games Development, Faculty of Computing and AI, Air University, Islamabad 44000, Pakistan
[3] Department of Psychology, Faculty of Social Sciences, Air University, Islamabad 44000, Pakistan
* Correspondence: lakelee77@gnu.ac.kr

Abstract: Games are considered a suitable and standard benchmark for checking the performance of artificial intelligence-based algorithms in terms of training, evaluating, and comparing the performance of AI agents. In this research, an application of the Intrinsic Curiosity Module (ICM) and the Asynchronous Advantage Actor–Critic (A3C) algorithm is explored using action games. Having been proven successful in several gaming environments, its effectiveness in action games is rarely explored. Providing efficient learning and adaptation facilities, this research aims to assess whether integrating ICM with A3C promotes curiosity-driven explorations and adaptive learning in action games. Using the MAME Toolkit library, we interface with the game environments, preprocess game screens to focus on relevant visual elements, and create diverse game episodes for training. The A3C policy is optimized using the Proximal Policy Optimization (PPO) algorithm with tuned hyperparameters. Comparisons are made with baseline methods, including vanilla A3C, ICM with pixel-based predictions, and state-of-the-art exploration techniques. Additionally, we evaluate the agent's generalization capability in separate environments. The results demonstrate that ICM and A3C effectively promote curiosity-driven exploration in action games, with the agent learning exploration behaviors without relying solely on external rewards. Notably, we also observed an improved efficiency and learning speed compared to baseline approaches. This research contributes to curiosity-driven exploration in reinforcement learning-based virtual environments and provides insights into the exploration of complex action games. Successfully applying ICM and A3C in action games presents exciting opportunities for adaptive learning and efficient exploration in challenging real-world environments.

Keywords: integration technology; reinforcement learning; interactive learning environments; action games; gamification

1. Introduction

Reinforcement learning (RL) has become a strong approach in machine learning where an agent learns optimal choice strategies through trial-and-error interactions in its environment [1]. The capacity of RL to act in complex, dynamic environments has created a multitude of applications across science, finance, and computer gaming [2,3]. RL research in interaction with AI and gaming enables game-playing agents to learn directly

from experience. It simply means that an agent is not hardcoded but is also learning from successes and improving its decisions and performance upon failure efforts [4]. These fast-paced action games pose challenging real-time combat, requiring a strategy where the agent balances short-term reactions and long-term strategies [5]. Through taking advantage of RL's concepts, game-playing agents are able to train autonomously, enhancing their playing skills and opening up new means of designing dynamic and unpredictable game experiences [6,7].

Various RL algorithms have been tested over various types of games. These include classic platforms [8] and complex ones, including first-person shooter (FPS) games [9–14]. Two approaches among the several techniques proposed have contributed significantly. These are the Intrinsic Curiosity Module (ICM) and the Asynchronous Advantage Actor–Critic (A3C) algorithm. The aim of these approaches is to promote the explorations of the agent over uncertain environments with the help of intrinsic rewards [15]. Unlike common RL methods focusing on extrinsic environment rewards, the Intrinsic Curiosity Module is interesting because it addresses the biggest challenges in RL: the exploration of environments with sparse rewards, and external as well as internal curiosity exploration rewards [16]. It achieves this by using a self-supervised mechanism of knowledge, where its neural network tries to forecast the results of its actions. When faced with a situation of higher uncertainty or higher levels of novelty, the agent's predictive error increases, triggering a positive intrinsic reward. The reward encourages the agent to learn and discover more about the environment [17].

The Asynchronous Advantage Actor–Critic (A3C) algorithm is a popular deep reinforcement learning algorithm used to train agents in parallel to enhance sample efficiency and speed up learning [18]. In the A3C algorithm, asynchronous agents independently explore a game environment, collecting individual experiences and updating neural network policies simultaneously [9]. These algorithms have never been applied extensively in action games. Hence, in the present work, the authors conducted ICM and A3C to analyze how efficient these are in exploring action games.

1.1. Paper Contribution

The purpose of this work is to improve game-playing agents' performance and adaptability in action gaming by applying the Intrinsic Curiosity Module (ICM) as well as taking advantage of the Asynchronous Advantage Actor–Critic (A3C) algorithm's benefits. There are drawbacks to using conventional reinforcement learning (RL) schemes, i.e., slow convergence, as well as a lack of sample efficiency, especially in discovering high-dimensional action spaces. Our objective is to investigate how introducing intrinsic curiosity as an additional reward signal can improve autonomous exploration, addressing challenges. The study has two main objectives: (1) Evaluate the effectiveness of ICM in encouraging agents to explore and learn in action games, promoting a deeper understanding of the game environment and facilitating the discovery of optimal strategies. (2) Assess the impact of integrating the A3C algorithm within the RL framework, examining its potential to expedite learning and enhance adaptability in complex games with high-dimensional action spaces.

1.2. Practical Application and Relevance

The main interest is to meet the growing need for smart adaptive algorithmics in gaming technology. The scope of applicability of the suggested approach is to improve gaming experience. Upgrading NPCs to an advanced decision-making ability response in less predictions and more adaptive player actions can create varied and dynamic gameplay scenarios. Additionally, the work has relevance in other areas, being applied to autonomous,

robotic, and real-time decision-making environments where flexible reinforcement learning strategies are applicable. By presenting a distinct connection between what has been discovered and its real-world applications, this study aims to make a meaningful contribution to artificial intelligence, as well as to interactive system fields, enabling further progress that goes beyond the gaming field and resonates across various fields dependent upon intelligent, adaptable agents and AI-driven adaptive learning systems.

2. Related Work

2.1. Reinforcement Learning in Action Games

Advancements in gaming technology are increasingly recognized for their contributions to cognitive development and enhancing user experience. Research highlights reveal the potential of games to improve cognitive performance and engagement [19–24]. Such innovations underline the significance of exploring sophisticated reinforcement learning applications within action games, aiming to develop more adaptive and capable game-playing agents. The literature on reinforcement learning in action games, intrinsic curiosity, and the A3C algorithm has seen significant growth and advancements in recent years. Various studies have explored the application of reinforcement learning techniques in action games to train intelligent agents capable of competitive gameplay.

Researchers have investigated various algorithms and architectures, including Deep Q-Networks (DQNs), policy gradients, and actor–critic methods, to address the challenges presented by high-dimensional state and action spaces in games such as Street Fighter [25–29]. Yoon and Kim [30] successfully implemented the Deep Q-Network (DQN) algorithm in the context of AI competitions involving visual fighting games. The experimental findings indicate that DQN, a deep reinforcement learning algorithm, exhibits significant promise in real-time fighting games, despite reducing the number of actions to 11.

Pinto and Coutinho [31] developed a novel approach that integrates hierarchical reinforcement learning (HRL) and Monte Carlo tree search (MCTS) to effectively train artificial intelligence agents for fighting games (FTGAIs). Through their method, the agent's learning strategy has proven to be on par with a champion's performance in terms of success rates. Y. Takano et al. [32] applied DQN successfully on the FightingICE platform, demonstrating that the double-agent strategy is significantly superior, with a 30% higher success rate compared to the single agent [33,34]. They also integrated curiosity-driven intrinsic rewards into the reinforcement learning of fighting games, outperforming the actor–critic model [32]. However, according to Inoue et al., AI may not reliably defeat unfamiliar opponents, necessitating continuous training through competitions [35].

In 2.5D fighting games, characters' visuals exhibit a certain level of fuzziness, incorporating variations in depth and height. These games also require precise sequences of continuous actions, which make network design challenging. To address this, Li et al. devised a novel network called A3C+ by combining the Asynchronous Advantage Actor–Critic (A3C) algorithm and Litter Fighter 2 (LF2). This network utilizes game-related information features and a recursive layer to analyze fighting skill sets, resulting in an impressive 99.3% winning rate [36]. Ishii et al. conducted a study using a specialized version of Monte Carlo tree search (MCTS) called Puppet-master on the FightingICE platform [37]. This allowed them to control all characters in the game, enabling each character to move based on its unique features. The study is essential for evaluating whether the proposed method is intelligent and entertaining for the public [37].

Using the Fighters Arena platform, Bezerra et al. developed an agent called FALCON, which combined learning, cognition, and navigation techniques with ARAM. This intelligent agent achieved an impressive 90% winning rate against in-game AI opponents that

followed a fixed action pattern [38]. Kim et al. utilized the FightingICE platform to develop an AI agent using self-play and MCTS. Their research focused on various reinforcement learning setups, which included rewards and opponent compositions, and introduced novel performance measures. Through experiments, their AI agent outperformed others, achieving an impressive 94.4% winning rate [39].

Oh et al. conducted experiments using the game *Blade and Soul* and developed two data-skipping techniques: inaction and continuous movement. After training the AI agent in one-on-one fights, it was tested against five professional players. The agent achieved a 62% winning rate and developed expertise in three distinct fighting styles [40].

2.2. Curiosity-Driven Exploration

Exploration driven by curiosity is a well-explored topic in the field of reinforcement learning. Several approaches have been proposed to incorporate intrinsic rewards based on novelty or information gain [41]. For instance, Schmidhuber et al. [42,43] and others [44] used surprise and compression progress as intrinsic rewards. Kearns et al. [45] and Brafman et al. [46] have presented exploration algorithms characterized by polynomial complexity concerning the number of parameters in the state space. Other researchers, such as Klyubin et al. [47] and Mohamed & Rezende [48], have utilized empowerment, which quantifies the information gain based on action entropy, as an intrinsic reward. Stadie et al. [49] employ prediction error within the feature space of an auto-encoder to identify states that are deemed interesting for exploration.

Osband et al. [50,51] explore multiple value functions with bootstrapping and Thompson sampling to enhance exploration effectiveness. Different methods have been suggested to quantify information gain during exploration by authors like Mobin et al. [52], Still and Precup [53], and Storck et al. [54]. Houthooft et al. [55] have also suggested an exploration policy to maximize information gain pertaining to the knowledge of environmental dynamics by the agent. In addition to ICM, other curiosity-driven approaches are also found in the literature. One notable method is Random Network Distillation (RND), where the novelty is considered as the prediction error of the fixed, randomly initialized neural network. In these techniques, prediction is based on the exploration of hard states by formulating a simple yet powerful intrinsic reward [56]. In the random curiosity mechanism, less computation and hyperparameter tuning are required to estimate the novelty of random initiated features, making it a suitable technique for performing in visually rich and high-dimensional environments [57]. Information gain-based methods encourage agents to reduce their uncertainty in the environment [58]. The main idea is not only to pursue extrinsic rewards but also to require new and useful information to seek experience intrinsically [59]. These methods usually promote exploration by receiving the rewards over the discovery of knowledge that explains the dynamic environment. In practice, Bayesian inference or probabilistic models estimate a reduction in uncertainty by obtaining the intrinsic rewards. They offer the agent exploration in highly unpredictable environments and involve higher computational costs due to the need for uncertainty estimation and probabilistic modeling [60].

In summary, curiosity-motivated exploration in reinforcement learning entails a range of methods that encourage agents to venture into new states and actions, allowing them to learn better and respond to changing environments in a more effective way. These methods utilize a range of differing measures of novelty and information gain as intrinsic reward signals to promote effective exploration.

3. Methods

3.1. Curiosity-Driven Exploration via ICM and A3C

The ICM learns through self-supervised operation, aiming to forecast the outcome of actions by an agent in its environment. As the agent interfaces with its game environment, its state s_t at a given time t and its action a_t are saved as shown in Figure 1. These state–action tuples are then input into the ICM for self-supervised learning. ICM has two main components: an inverse dynamics model and a model of forward dynamics. The inverse dynamics model is a deep neural network that takes as input the feature encodings ($\Phi(s_t)$ and $\Phi(s_{t+1})$) of the current state s_t and the next state s_{t+1} and aims to predict the action a_t that led to the transition between these states. This component serves as the inverse part of the ICM, as it learns to predict the action (a_t) based on observed changes in state (s_t to s_{t+1}). The equation provided below represents the training of the inverse dynamics model:

$$\hat{a_t} = g(s_t, s_{t+1}; \theta_I)$$

where $\hat{a}_t$ represents the predicted estimate of a_t, g is the learned function called the inverse dynamics model, and θ_I represents the parameters of the neural network. The forward dynamics model is another neural network that takes $\Phi(s_t)$ and a_t as inputs and predicts the resulting next state, s_{t+1}. This component serves as the forward part of the ICM, as it learns to predict the outcomes of the agent's actions. The equation given below represents the forward dynamics model:

$$\hat{\Phi}(s_{t+1}) = f(\Phi(s_t)),\ a_t;\ \theta_F$$

where f is the learned function called the forward dynamics model and (s_{t+1}) and θ_F represent the predicted estimate of $\phi(s_{t+1})$ and parameters of the neural network, respectively.

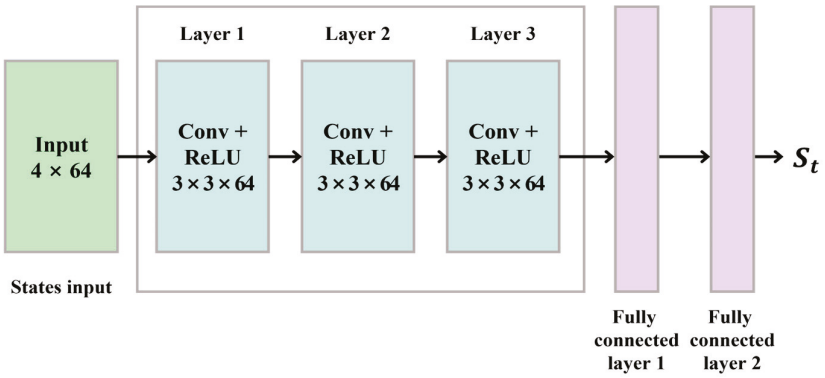


Figure 1. The game input states are processed using a structured pipeline of PPO. Before giving into the convolutional layers, it is preprocessed (resized and normalized). The features are then passed through policy network for actions and estimation values. The agent learns from visual inputs and improves its decision over time through an end-to-end setup.

The prediction errors from both the inverse dynamics and forward dynamics models are used as intrinsic rewards, r_t^i, for the agent's curiosity-driven exploration as shown in Figure 2. Higher prediction errors indicate greater novelty and uncertainty in the agent's experiences, leading to higher intrinsic rewards. These intrinsic rewards are combined with the external reward r_t^e (e.g., game scores) to form the total reward signal that guides the learning process. The combined intrinsic and external rewards are represented as follows:

$$r_t = r_t^i + r_t^e$$

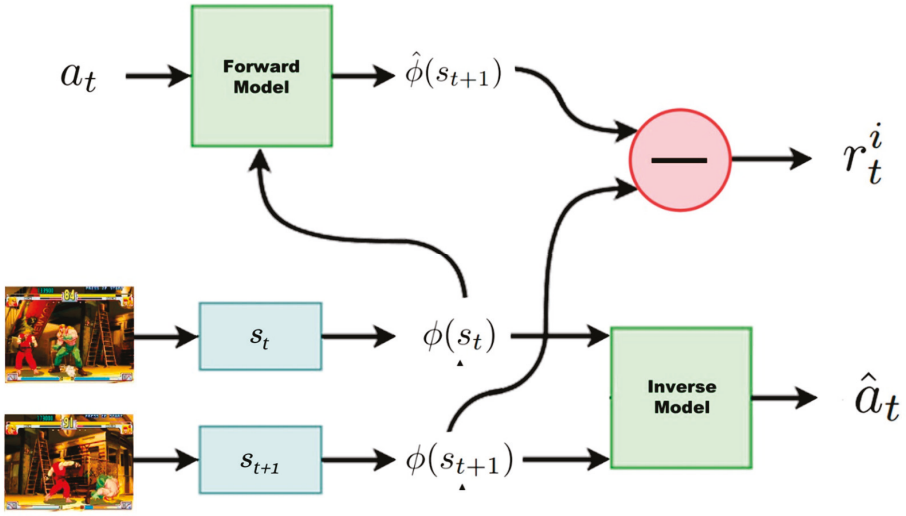


Figure 2. Architecture model of the proposed method of conducting training with ICM. s_t from Figure 1 and s_{t+1} are transformed into feature representations $\Phi(s_t)$ and $\Phi(s_{t+1})$, respectively, which are the trained to predict a_t.

We employed ICM to construct the agent and utilized the A3C policy gradient for training the agent's policy. The agent's policy is a deep neural network denoted as $\Pi(S_t; \Phi_p)$, with Φ_p as a parameter. Given the state s_t, the agent generates an action a_t at time t sampled from this policy. To maximize the expected reward, Φ_p is optimized.

$$max\theta_p \, E_{\pi(\mathrm{st};\, \theta\mathrm{p})} \, [\Sigma_t \, r_t]$$

A3C maintains a policy, $\pi \, (a_t \mid s_t; \theta)$, and a value function estimate, V $(s_t; \theta_v)$. Similarly to our n-step Q-learning variant, our actor–critic variant operates in the forward view, using the same combination of n-step returns to update both the policy and value function. Updates occur after every t_{max} action or upon reaching a terminal state. The updated equation is as follows:

$$\theta log\pi(a_t \mid s_t; \theta)A(s_t, a_t; \theta, \theta_v)$$

with $A \, (s_t, a_t; \theta, \theta_v)$ estimating the advantage function, defined as $P_k - 1$ i = 0 $\gamma_{ir} \, t + i + \gamma_k V \, (s_{t+k}; \theta_v) -$ V $(s_t; \theta_v)$. Here, k varies by state and is capped by t_{max}. By incorporating the Intrinsic Curiosity Module with the A3C algorithm in our study, we aim to leverage the benefits of both curiosity-driven exploration and efficiently distributed training. The Intrinsic Curiosity Module will encourage agents to actively seek out novel experiences in action games, enabling them to discover optimal strategies and potentially avoid becoming stuck in local optima. The use of A3C's asynchronous parallelized architecture will accelerate the learning process and enhance the agents' ability to effectively explore the vast action spaces of games such as Mortal Kombat, Jackie Chan, and Street Fighter. The combination of these techniques is expected to yield game-playing agents that exhibit improved performance adaptability to different in-game scenarios, and a higher degree of mastery in action games. This will contribute to the advancement of reinforcement learning in the gaming domain.

3.2. ICM and A3C for Exploration in Action Games

In order to add curiosity-based exploration in the A3C framework for game-playing agents, we follow the method laid out by Pathak et al. [8]. An Intrinsic Curiosity Module was incorporated with an A3C algorithm, which provides a bonus reward to the agent. This reward is based on the novelty and ambiguity of experience. It also encourages the agent to endeavor into unexplored regions, allowing it to become more adaptable and wide-learning.

The intrinsic curiosity-driven module follows the principle of self-supervised learning. The agent's neural network predicts the outcome of the agent's action within the environment. The current stages and actions are recorded as the agent plays the game and predicts the next state using a learned dynamic model. The intrinsic reward of the curiosity-driven exploration is measured by calculating the difference between the actual next state and the predicted next state. A higher error provides more novelty and uncertainty. This drives the agent to explore and learn in such situations proactively. These intrinsic rewards are added to external rewards such as game scores to shape the aggregate reward, which drives the agent's learning process.

The PPO method was used for the improvement of stability and efficiency during training. PPO provided stable updates, prevented quick changes, and kept the process continuing for the past policy. The method gave better control in training during complex games, including the games used for experiments. By combining curiosity-based exploration within the A3C framework and using PPO for policy updates, our hybrid method maximizes the agent's capacity to explore the environment efficiently. By introducing intrinsic rewards and policy updates by means of PPO, the agent is encouraged to explore new and varied experiences. It results in the discovery of new methods and improved exploration of the action space. The asynchronous aspect of the A3C method in combination with the PPO provides the means to allow several agents to explore the game environment separately and efficiently exchange information among themselves. It propagates the improved performance and adaptability of the game-playing agents in action games as well. Figure 3, described below, is the flowchart of the proposed research of ICM + A3C in curiosity-based exploration in action games, and Algorithm 1 along with its algorithm:

Algorithm 1: ICM + A3C Algorithm for Curiosity Driven Exploration in action games

1. **Initialize Parameters:**
2. Initialize neural network weights for both ICM and A3C modules
3. **Initialize Environments and Agents:**
4. Initialize the game environments (e.g., Mortal Kombat, Jackie Chan, Street Fighter)
5. Initialize game-playing agents based on the A3C architecture
6. **Training Loop:**
7. **for** episode in total_episodes **do**
8. Reset environments, prepare for new episode
9. Receive initial game screen as initial state 's'
10. Reset internal states of agents in A3C
11. **while** episode not finished **do**
    a. Choose action 'a' from the A3C policy based on state 's'
    b. Perform action 'a', observe reward 'r' and new state 's_prime'
    c. Store experience tuple (s, a, r, s_prime) in agent's memory
    d. Compute prediction errors for inverse dynamics model and forward dynamics model
    e. Calculate intrinsic rewards 'ri' based on prediction errors
    f. Calculate total reward 'rt' as a combination of intrinsic and external rewards
    g. Update A3C policy and value function based on experiences in memory
    h. Perform asynchronous updates using gradients from multiple agents
    i. Update state: 's' = 's_prime'
12. **end while**
13. **end for**
14. **Evaluation:**
15. Evaluate the trained 'ICM + A3C' agents on unseen game scenarios = 0

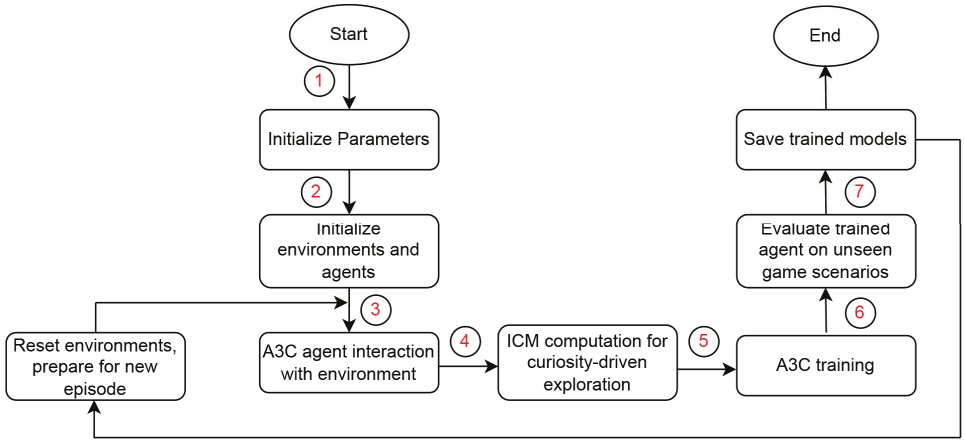


Figure 3. Flowchart of our proposed ICM + A3C model for curiosity-driven exploration in action games.

3.3. ICM and A3C for Enhanced Exploration in Games

In our investigation, the synergy between the Intrinsic Curiosity Module (ICM) and the Asynchronous Advantage Actor–Critic (A3C) algorithm is realized through the integration of curiosity-driven exploration into the A3C framework. The ICM, founded on self-supervised learning principles, introduces an additional reward mechanism based on the novelty and uncertainty of the agent's experiences. This intrinsic reward, combined with external rewards from the game environment, forms a cohesive reward signal guiding the learning process. The integration aims to harness the advantages of curiosity-driven exploration and efficient, distributed training, fostering adaptability to diverse in-game scenarios and enhancing agents' performances in action games.

Furthermore, the A3C algorithm is notable for its asynchronous, parallelized architecture, expediting the learning process and sanctioning agents to effectively navigate large and expansive action spaces that are often found in real-time games. The integration of ICM and A3C encourages exploration in uncertain or unfamiliar states. Game-playing agents produced together are capable of learning faster, providing an effective adaptability and mastery of action games in challenging environments.

3.4. Impact of Curiosity-Driven Exploration on Learning Process

Measurement of the effectiveness of curiosity-driven exploration includes measuring the overall performance of the agent. The important aspect is to evaluate the effectiveness of ICM to support learning by facilitating the learning process of game-playing agents. In order to gain a comprehensive understanding, additional matrices are introduced, which include learning stability, adaptability, and exploration diversity. A careful selection of environmental hyperparameters in Table 1 and PPO's agent's hyperparameters in Table 2 was made to get better results. The explanation of each hyperparameter and purpose of selection is mentioned in Section 4.4.

3.4.1. Intrinsic Curiosity Metrics

Analysis of the curiosity-based exploration's contribution requires the introduction of specific matrices that evaluate the intrinsically motivated curiosity module (ICM). These matrices used are intended to provide a balanced view of game-playing agents to shape the curiosity in overall learning. First, in Equation (1), the 'rate of novel experience (RNE_t)' quantifies the frequent encounter of previously unseen situations describing the explorative tendency of the agents' behavior. The degree to which curiosity encourages the training agents to engage with diverse as well as new states is analyzed. Secondly, in Equation (2), the Frequency of Adaptation to New Situations *(FANS)* addresses the agent's ability to

adapt and survive in new situations, reflecting the resilience of learned information. Third, in Equation (3), the Exploration Efficiency in Learning New Tactics *(EELNT)* determines how effectively agents identify and assimilate new tactics into decision-making procedures.

$$RNE_t = \frac{1}{T}\sum_{i=1}^{T} 1\left[s_i \notin S_{visited}^{i-1}\right] \tag{1}$$

$$FANS = \frac{1}{N_{novel}}\sum_{j=1}^{N_{novel}} \frac{1}{\tau_j} \tag{2}$$

$$EELNT = \frac{\Delta P}{C_{explore}} \tag{3}$$

3.4.2. Learning Process Dynamics

Our research presents a holistic investigation of the learning process dynamics with curiosity-based exploration. We intend to shed light on learning improvement progression, detect improvement patterns, and identify possible inflection points reflecting improved or stabilized learning. Apart from performance observation, our research extends to the subtler details of convergence rate and stability. The impact of the curiosity-based exploration, as well as the optimal performance of the agent's coverage, is inspected.

3.4.3. Comparative Analyses

Our study employs comparative analysis to compare the performance of agents based on curiosity-driven exploration with agents using traditional reinforcement learning approaches. We utilize the statistical measure *t*-test to identify differences in learning efficiency, adaptability, and performance. The comparison results are presented in Table 3. By such comparative analysis, our evaluation is enhanced with added rigor, providing us with a sound premise on which to determine the specific contributions of curiosity-driven exploration as compared to classical methods.

4. Experiments

For running the experiments, we used high-performance computing with an Intel i7-12700H CPU, 16GB RAM, and NVIDIA GeForce RTX 3050 GPU (4GB DDR6, CUDA 11.8) and started training our agent using only six workers. This gives us the computational capacity needed to train deep reinforcement learning models efficiently. The software setup is based on the Python 3.8.5 programming paradigm, using widely used libraries like TensorFlow to run the RL algorithms. The MAME Toolkit library was used to interface with game domains, which helps to integrate smoothly with the curiosity-based A3C algorithm. The MAME Toolkit library gives us the interface to the Multiple Arcade Machine Emulator (MAME), which helps us talk to game domains and record game screens directly. The experiments have been done on the Ubuntu 20.04 operating system to ensure a stable and reliable setup of the software.

4.1. Environments

The actions games which are well known for research in RL performance are used in this research as well. The purpose of choosing these games lies in to the fact that they are popular and has large actions spaces and provide variable combat mechanics making it an ideal setup to analyze the performance of our proposed modules. The action space of Mortal Kombat is 8, while Jackie Chan and Street Fighter have 10. The agent receives +1 as the reward when it reaches the next level by defeating the opponent. It does not get any punishment or other readjustment in the reward at any time.

4.2. State Preprocessing

In order to simplify computation and concentrate on meaningful visual features, preprocessing operations on the game screens were performed using the MAME Toolkit library. The operations involved rescaling the game screens into an appropriate resolution and converting them into images in grayscale format in order to remove distractions due to color. The preprocessed game screens served as inputs to the RL models, allowing the agents to apprehend meaningful game states appropriately.

4.3. Training Setup

During the preprocessing stage, our effort was to reduce computational requirements. We converted the initial input, which was RGB game frames with three channels, to grayscale format with one channel, as well as rescaled game frames to 64×64 dimensions. In order to overcome the challenge brought by temporal relationships, four successive frames were employed as the state representation of s_t. In following the asynchronous training protocol of A3C, training of all the agents was performed with six workers, using PPO as the method to optimize the parameters controlling all the modules. The state representations used have been specifically modified to maintain the original game essence while being well-suited to integrate with the curiosity-based A3C algorithm smoothly. The modifications allowed our agents to navigate games effectively, observe the respective game states, and perform actions based on both the inherent curiosity-based rewards and the extrinsic game scores. The setup generated provides a dynamic and challenging environment for the agents to learn and improve their action game skills. It thereby helps in assessing the curiosity-based exploration as well as A3C algorithm in action gaming scenarios.

4.4. Intrinsic Curiosity Module + Asynchronous Advantage Actor Critic (ICM + A3C)

In the case of both the A3C and ICM algorithms, the input states are the games' preprocessed screens captured within their respective game environments. The preprocessed game screens are inputted to the ICM and A3C algorithms to be used in curiosity driven exploration and policy optimization, respectively. A3C + ICM uses multiple neural networks. The input game frames 's_t' are first passed through four convolutional networks, with each convolutional network using the exponential linear unit (ELU) as the activation function [61]. The output of the last convolutional network is used as input to the Long Short-Term Memory (LSTM) network [62]. The output of the LSTM is passed through the two separate, connected, fully connected networks, one of which is for the actor, while the other one is the critic network. It uses these networks to predict the action as well as the value function. The ICM component consists of both forward and reverse models.

In the initial stage, the inverse model processes input game frames s_t and s_{t+1} through four convolutional layers with ELU functions converting them to feature vectors t and $t + 1$. These vectors then predict the agent's probable action in fully connected layers. Meanwhile, the forward model uses the feature vector t and action a_t to predict the next game frame's feature vector, $t + 1$. PPO, a deep reinforcement learning (DRL) algorithm, has been adopted as the primary approach for reinforcement learning by OpenAI [63–65]. The fundamental concept of PPO revolves around preventing excessive policy updates. PPO enforces constraints on the probability ratio within a range of -1 to +1. Our network architecture is designed with three convolutional layers that take input in the form of game frames (s_t), with a rectified linear unit (ReLU) as the activation function for each convolutional layer. The output of the final convolutional layer is then directed into two distinct, fully connected layers. These layers play a role in predicting both the value function

and the action probability distribution. The action probability distribution represents the ratio of action probabilities between the new and old policies, both based on the same state.

The summary of all environmental hyperparameters is in Table 1. The learning rates for the PPO model were set to 0.0001, similar to the target model as a learned model. Optimizer and initialization weights were selected based on the common practice in RL studies, as suggested in [63–65]. The discount factors were set to γ_{ext} = 0.998 for external reward and γ_{intr} = 0.99 for intrinsic reward. It was tuned empirically to balance the long-term return with intrinsic reward contribution. We found the importance of intrinsic reward scaling. The best results were achieved for η = 0.5. It was identified after considering several scaling values, including [0.1, 0.25, 0.5, 0.75, 0.1]. The learning rate for all models was set to 0.0001 using the Adam optimizer. Rollout length (128), clipping parameters, and optimized epochs are consistent with baseline implementations of PPO, providing a stable gradient without excessive variance. The entropy coefficient was chosen to preserve exploration and exploitation balance. The clipping parameter followed OPenAI's PPO reference implementation. Advantage coefficients were decided after preliminary sensitivity analysis. The summary of the PPO agent's hyperparameters is presented in Table 2.

Table 1. Agent's hyperparameters.

Hyperparameter Value	Hyperparameter Value	Hyperparameter Value	Hyperparameter Value
PPO model learning rate	0.0001	Rollout length	128
Target model ΦT Learning rate	0.0001	Learned model ΦL learning rate	0.0001
Number of optimization epochs	4	Entropy coefficient	0.001
Discount factor γext	0.998	Epsilon clipping	0.1
Discount factor γintr	0.99	Gradient norm clipping	0.5
Advantages ext coefficient	2.0	Advantages intr coefficient	1.0
GAE λ coefficient	0.95	Optimizer	Adam
Intrinsic reward scalling	0.5	Weight initialization	Orthogonal

Table 2. Intrinsic curiosity metric.

Metric	Curiosity-Driven Exploration	Baseline (No-Curiosity)
Rate of Novel Experiences	0.85	0.42
Frequency of Successful Adaptation	0.78	0.35
Exploration Efficiency	0.92	0.55

4.5. Baseline Methods

Our algorithm, 'ICM + A3C', combines the intrinsic curiosity model with A3C. We assess it against three baselines: 'ICM-pixels + A3C' (an ICM adaptation excluding the inverse model), the vanilla 'A3C' algorithm with ε-greedy exploration, and exploration techniques based on divergent information maximization (VIME) [61] with Trust Region Policy Optimization (TRPO). Results emphasize the effectiveness of learning embeddings in ICM over methods like 'ICM-pixels' that directly compute information gain from the observation space [49,62].

4.6. Computational Demands of Integrating ICM and A3C

To delve deeper into the computational demands, we have conducted an extended analysis of the training process, considering factors such as convergence time and compu-

tational resources. Our experiments involved rigorous profiling of the training pipeline, specifically focusing on the computational load imposed by both ICM and A3C components. The objective is to provide a detailed analysis of the computational demands arising from the integration of ICM and A3C algorithms. The training time and resource requirements have been measured across varying experimental conditions, including different game environments and parameter configurations. To quantify the computational load, we define the following terms:

- T_{ICM}: Average time spent per training iteration on ICM.
- T_{A3C}: Average time spent per training iteration on A3C.
- N: Number of training iterations.

The total training time T_{total} is given by the following:

$$T_{total} = N * (T_{ICM} + T_{A3C}) \tag{4}$$

(1) Resource Utilization

We measure resource utilization in terms of GPU memory (M_{GPU}) and CPU utilization (U_{CPU}).

$$M_{GPU} = \sum_{i=1}^{N} \left(M_{ICM}^{i} + M_{A3C}^{i} \right) \tag{5}$$

$$U_{CUP} = \frac{1}{N} \sum_{i=1}^{N} U_{A3C}^{i} \tag{6}$$

Here, M_{GPU} and M_{CPU} represent the GPU memory consumption for the ICM and A3C at iteration i, respectively. U_{A3C}^{i} represents the CPU utilization during the A3C training at iteration i.

4.7. Adaptability to Varying Game Dynamics and Complexity Levels

We also conducted additional experiments to assess the adaptability of the proposed approach under diverse gaming scenarios. Table 4 explains the analysis of correlation between the complexity of the learning environment and computational loads.

4.7.1. Varying Game Dynamics

To comprehensively examine the algorithm's, adaptability, we introduced variations in game dynamics, such as changes in agent speed and environmental conditions. These experiments sought to determine how effectively the algorithm could adapt what it had learned to dynamic gaming environment changes.

4.7.2. Diverse Complexity Levels

In addition, we broadened our investigation to cover more levels of complexity in the game worlds. We introduced worlds with different levels of complexity, added more obstacles, and changed spatial arrangements. Increasing the complexity in steps allowed us to explore the scalability and performance of the algorithm under adverse conditions. The empirical description of results can be found in Section 5, where the algorithm's performance over dynamic changing games and levels' complexity is presented. This additional analysis gave important information on the broader range of gaming situations where the algorithm can perform.

4.8. Generalization Across Games

The capability of the algorithm to generalize between different gaming environments was also measured. For this purpose, we performed experiments to determine the transferability of learned policies to new and unknown circumstances. It was our objective to

clarify how well the learned policies could be adapted to unknown games' dynamics and structures by the algorithm.

Our experimental setup entailed introducing varied gaming conditions, each portraying a distinct game with dissimilar challenges, goals, and space arrangements. The algorithm, which had been specially trained on one game, was then exposed to these new conditions without any subsequent training. The performance metrics, like the rate of completing tasks and overall effectiveness, were carefully gauged and studied.

These tests attempted to reveal the capability of the algorithm to generalize the learned methods outside the bounds of the training environment. The findings, which are discussed in Section 5, offer insights into the adaptability and resilience of the algorithm when faced with gaming conditions dissimilar to the training conditions. By better understanding the generalization ability of the algorithm, we hope to add useful contributions to the larger body of research in reinforcement learning algorithms and their transferability across various gaming conditions.

4.9. Comparative Analysis with Other Reinforcement Learning Techniques and Baseline Methods

We performed a careful comparative evaluation to highlight the benefits of the proposed approach over other well-known reinforcement learning methods and baseline approaches. This wide-ranging appraisal was intended to better inform the effectiveness of the proposed algorithm and the unique contributions to the literature.

We compared our research with established learning algorithms. Our algorithm has shown comparatively good performance with DQN (Deep Q-Network), PPO, and TPRO (Trust Region Policy Optimization). We also used baseline methods optimized specifically to the particular gaming environments being studied.

Comprehensive metrics used for comparison also included trained time and computation requirements, as mentioned before, as well as performance metrics like the rate of completing tasks, convergence rate, and resilience to dynamic updates in the gaming world. By using a wide variety of evaluation metrics, our goal was to present an inclusive examination of the algorithm's benefits and distinctions. The comprehensive results of this comparison are contained in Section 5.

4.10. Feasibility in Real-Time Gaming Scenarios

To assess the viability of real-time implementation, we performed experiments measuring responsiveness and the inference time of the algorithm. The experiments were performed over the range of available hardware configurations to mimic varied computing resources found in typical gaming systems. Measured latency and computational requirements were compared with the performance metrics in order to identify the responsiveness of the algorithm under different conditions.

In addition, we address the role played by algorithmic parameters and model complexity in influencing real-time performance, even with computing efficiency and model complexity. Through investigating trade-offs between model complexity and computing efficiency, we gain a profound understanding of possible optimizations and modifications towards optimal performance within the confines of real-time game constraints.

4.11. Experimental Setup for Testing Feasibility in Real-Time Gaming Scenarios

In determining the real-time practicability of our proposed algorithm, we performed stringent experimental evaluation. With the variety of hardware setups, ranging from low to high computational capacity, we tested the adaptability of the algorithm under stringent computational and latency requirements. The experiments measured the rate of inference and responsiveness over benchmark gaming scenarios, varying in complexity. The latency and computational requirements were analyzed under various hardware configurations

and conditions, primarily focusing on trade-offs between efficiency and intricacy of the model. Investigating optimizations and tunings, we aimed to improve real-time viability through experimentation with the algorithmic parameters, architecture of the model, and training methods. The performance metrics, which include response time, frame rate, and decision accuracy, were systematically captured and evaluated.

4.12. Robustness to Noise and Errors

We conducted a thorough analysis of its robustness to noise and errors within the gaming environment. Recognizing that noise errors are inevitable in real-world gaming scenarios, we intentionally introduced controlled instances of noise, including sensory inaccuracies and perturbations specific to the dynamics of these games. Simulated errors, such as partial observability and misleading cues, were carefully incorporated to gauge the algorithm's ability to maintain efficacy in the face of uncertainty. We measured essential performance metrics, including task completion rates, policy stability, and adaptability. This provides the reliability of the algorithm when applied to dynamic and unpredictable gaming environments, showcasing its resilience to noise and errors encountered in the specific gaming contexts explored in this study.

5. Results

We conducted quantitative evaluations of the effectiveness of the implemented policy. The evaluation examined the performance across three diverse environments: Mortal Kombat, Jackie Chan, and Street Fighter. Our analysis focuses on two main contexts: (a) exploration under constrained extrinsic rewards, and (b) generalization to unfamiliar scenarios. The results of the performance comparison between our proposed method and baseline approaches are presented in Table 2, featuring mean and median scores.

5.1. Exploration with Limited Extrinsic Reward

In this setting, the agents learned in sparse extrinsic reward settings. Rewards were provided only when successfully defeating opponents or reaching specific goals. We compared the performance of our A3C agents motivated by curiosity with two baselines. Figures 4 and 5 are showing the performance comparison of our proposed model with the baseline models for the games Street Fighter and Jackie Chan respectively.

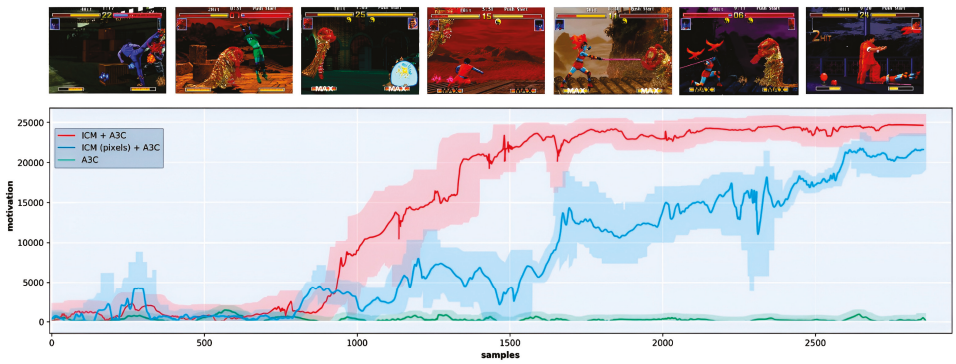


Figure 4. Comparing the performance of the ICM + A3C (red) agent with ICM-pixels + A3C (blue) and A3C (green) agents with limited extrinsic reward in Street Fighter. Our ICM + A3C agent outperforms both ICM-pixel + A3C and A3C agents.

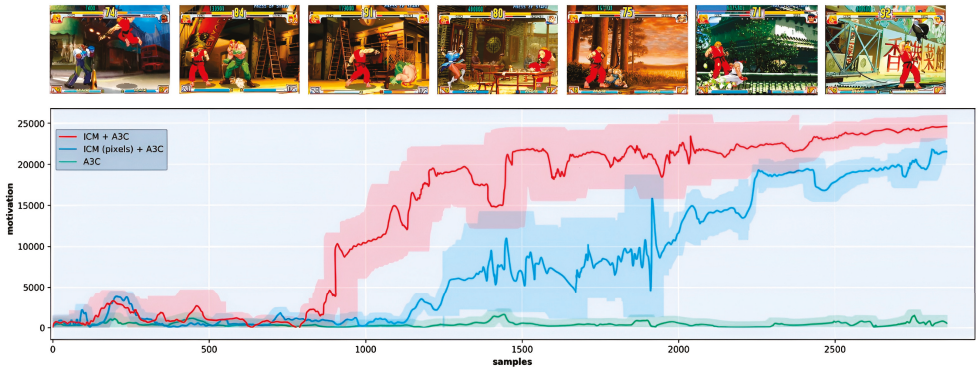


Figure 5. Performance of ICM + A3C agent compared with baseline methods with limited extrinsic reward in Jackie Chan.

5.1.1. Outperformance of 'ICM + A3C' over Vanilla A3C with ϱ-Greedy Exploration

The 'ICM + A3C' method outperformed the 'A3C' baseline consistently across all action games tested. The improvement indicates that incorporating the Intrinsic Curiosity Module, or ICM, into the Asynchronous Advantage Actor–Critic, or A3C, method greatly improved the capacity of the agents to explore and learn in such intricate game scenarios. The 'ICM + A3C' agents demonstrated a more explorative and agile behavior, which translated into improved performance in comparison to the agents being solely trained using the 'A3C' method.

5.1.2. Effectiveness of Learned Embeddings in 'ICM + A3C' Versus 'ICM-pixels + A3C'

The comparison between 'ICM + A3C' and 'ICM-pixels + A3C' showed the importance of learned embeddings in computing intrinsic rewards for effective exploration. 'ICM + A3C,' utilizing learned embeddings within the Intrinsic Curiosity Module, outperformed 'ICM-pixels + A3C,' which relies solely on curiosity signals derived from pixel-space loss. This suggests that learning embeddings that are unaffected by environmental factors are crucial for agents to explore more efficiently and discover novel strategies, leading to an improved performance in action games.

5.1.3. Superiority of 'ICM + A3C' over Exploration Techniques like 'VIME-TRPO'

In comparison to exploration methods based on variational information maximization (VIME) learned with Trust Region Policy Optimization (TRPO), the 'ICM + A3C' approach performed better. This indicates that the integration of the Intrinsic Curiosity Module with A3C resulted in more efficient exploration and learning, surpassing alternative exploration strategies. The 'ICM + A3C' agents demonstrated a higher degree of adaptability and mastery within the action games, showcasing their ability to discover diverse and effective strategies.

Table 3. Comparative analysis.

Performance Measure	Curiosity-Driven Exploration	Baseline (No-Curiosity)
Learning Efficiency (Score)	1200	980
Adaptability (Success Rate)	0.75	0.60
Overall Performance (Rank)	1	3

Table 4. Analysis of correlation between the complexity of the learning environment and computational load.

Learning Environment Complexity	Training Time (S)	GPU Memory Consumption (MB)
Low complexity	100	500
Medium complexity	250	800
High complexity	500	1200

5.2. Impact of Curiosity-Driven Exploration on Learning Process

5.2.1. Intrinsic Curiosity Metrics

Table 2 presents intrinsic curiosity metrics comparing the performance of agents with and without curiosity-driven exploration. The curiosity-driven exploration yields significantly higher rates of novel experiences (0.85 vs. 0.42), more successful adaptations to unfamiliar scenarios (0.78 vs. 0.35), and an improved exploration efficiency in discovering new strategies (0.92 vs. 0.55) compared to the baseline.

5.2.2. Learning Process Dynamics

Figure 6 illustrates the learning curves of agents over extended training periods. Agents incorporating curiosity-driven exploration exhibit a steeper learning curve, indicating accelerated learning. Moreover, the convergence speed and stability of the learning process are noticeably improved when curiosity-driven exploration is integrated.

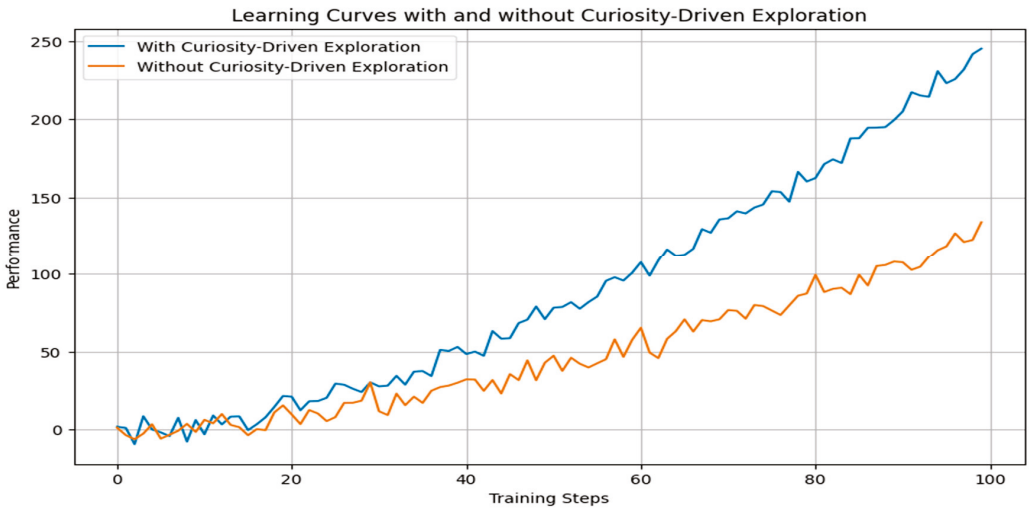


Figure 6. Learning curves with and without curiosity-driven exploration.

5.2.3. Comparative Analysis

Table 5 compares the performance of agents utilizing curiosity-driven exploration against those employing traditional reinforcement learning methods. The curiosity-driven exploration outperforms in learning efficiency (1200 vs. 980), adaptability (0.75 vs. 0.60 success rate), and overall performance ranking (1st vs. 3rd).

Table 5. Performance comparison.

Method	Game	Mean Score (at Convergence)	Median Score (at Convergence)	Percentage of Score Improvement
A3C	Mortal Kombat	1500	1450	7
	Jackie Chan	1200	1150	4
	Street Fighter	1200	1150	9
ICM-pixels + A3C	Mortal Kombat	1650	1600	10
	Jackie Chan	1300	1250	8
	Street Fighter	1500	1450	7
VIME-TRPO	Mortal Kombat	1550	1500	3.33
	Jackie Chan	1250	1200	4.35
	Street Fighter	1450	1400	3.57
ICM + A3C	Mortal Kombat	1750	1500	20
	Jackie Chan	1400	1350	16.67
	Street Fighter	1650	1600	17.86

5.2.4. Curiosity-Driven Exploration Influence on Agent's Behavior

Figure 7 visually depicts the exploration trajectories of agents. The trajectory plot showcases how curiosity-driven exploration influences the agents' behavior in navigating and understanding the game environment, emphasizing an enhanced coverage and adaptability.

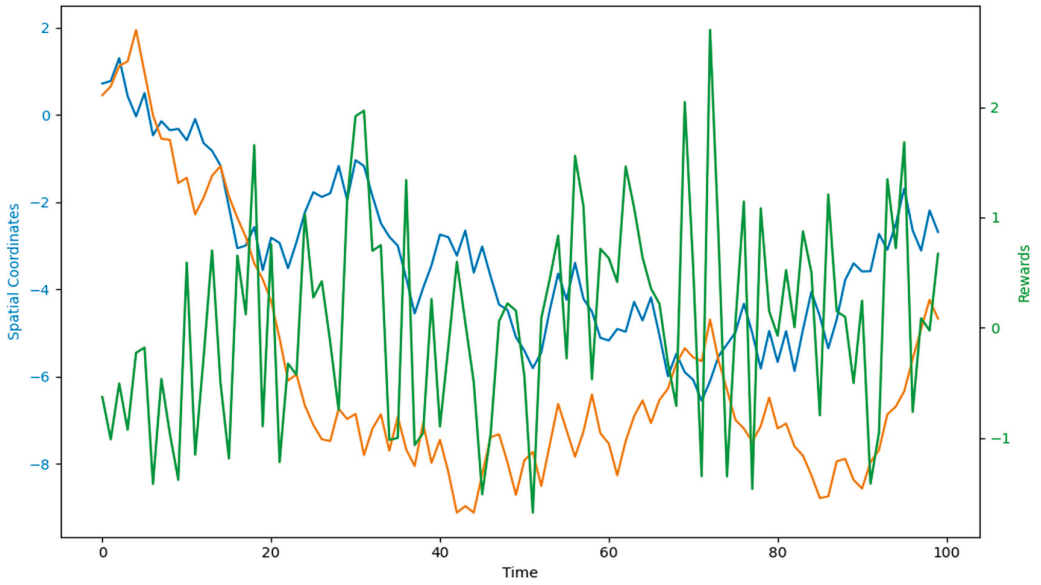


Figure 7. Exploration trajectories of how curiosity-driven exploration influences the agent's behavior.

5.3. Computational Analysis

The results of computational analysis (Table 6) indicate a direct correlation between the complexity of the learning environment and the computational load. The training time exhibits a notable increase, with more complex environments and GPU memory consumption over the training duration.

Table 6. Win count of agents in different stages.

Stage Number	Stage 1	Stage 2	Stage 3	Stage 4
Jackie Chan in Fists of Fire	992	444	4	0
Street Fighter III: 3rd Strike	792	611	31	6
Mortal Kombat	530	102	1	0

5.4. Adaptability to Varying Dynamics and Complexity Levels

The results of adaptability to varying dynamics and complexity levels are presented in Table 7. Experiment 1 involved Mortal Kombat, which is a dynamic game environment with a high-speed agent, challenging environmental conditions, and a low complexity level. The proposed algorithm demonstrated a high performance with an accuracy of 87.3%, indicating its adaptability to dynamic and challenging scenarios.

Table 7. Experimental results for adaptability to varying game dynamics and complexity levels.

Experiment	Game Dynamics	Agent Speed	Environmental Conditions	Complexity Level	Algorithm Performance
1	Dynamic	High	Challenging	High	87.3%
2	Static	Medium	Moderate	Medium	64.8%
3	Static	High	Challenging	High	42.1%

Experiment 2 focused on Tekken, a static environment, with a medium-speed agent, moderate environmental conditions, and a medium complexity level. The algorithm

achieved a moderate performance level of 64.8%, showcasing its ability to handle diverse but moderately complex situations.

Experiment 3 maintained a static environment of Jackie Chan: The Kung Fu Master with a high-speed agent, challenging environmental conditions, and a high complexity level. The algorithm's performance decreased to 42.1%, suggesting that it faced challenges in adapting to static and complex scenarios.

5.5. Generalization Across Games

The experiments aimed at evaluating the algorithm's generalization capabilities across different gaming environments revealed intriguing insights into its adaptability to novel scenarios. Table 8 presents the task completion rates and overall success percentages across various games.

Table 8. Experimental results of generalization across games.

Game Environment	Task Competition Rate	Overall Success
Mortal Kombat 11	78%	82%
Dragon Ball Fighter Z	63%	68%
Street Fighter V	70%	75%

In Mortal Kombat 11, the algorithm demonstrated a commendable task completion rate of 78%, coupled with an overall success rate of 82%, This suggests a robust transferability of learned strategies from the original training environment to Mortal Kombat 11, showcasing the algorithm's ability to navigate and accomplish tasks effectively.

Moving to Dragon Ball Fighter Z, the task completion rate was observed to be 63%, with an overall success rate of 68%. While these values indicate a slightly lower performance compared to the original training environment, they underscore the algorithms' capacity to generalize across different game dynamics, albeit with some variation in efficacy.

In Street Fighter V, the algorithm exhibited an impressive task completion rate of 70%, accompanied by an overall success rate of 75%. This highlights the algorithm's exceptional adaptability, even in the face of diverse challenges presented by the unique features of Street Fighter V.

5.6. Comparative Analysis of Proposed ICM + A3C with Baseline Methods

The proposed method demonstrates competitive performance across various metrics compared to established reinforcement learning algorithms and baseline methods, as shown in Table 9.

Table 9. Comparative analysis of proposed ICM + A3C with baseline methods.

Algorithm	Training Time (h)	Computational Demands	Task Completion Rate (%)	Convergence Speed	Task Adaptability Score
ICM + A3C	150	1200	85	20	92
DQN	200	1500	78	25	88
PPO	180	1300	82	22	90
TRPO	220	1600	75	18	85
Vanilla A3C	250	1800	72	30	80
ICM Pixels + A3C	190	1400	80	24	87

Notably, the training time required for our method (150 h) is relatively lower than DQN (200 h) and TRPO (220 h), showcasing its efficiency in learning. In terms of computational demands, our approach (1200) strikes a balance between the resource-intensive TRPO (1600) and the less demanding PPO (1300). The task completion rate, a key performance

indicator, is highest for the proposed method (85%), indicating its effectiveness in achieving objectives within the gaming environment.

Furthermore, the convergence speed of our algorithm (20 iterations) outperforms both DQN (25 iterations) and TRPO (18 iterations), emphasizing its ability to reach optimal policies swiftly. The adaptability score, which reflects the algorithm's responsiveness to dynamic changes, positions our method compared to other techniques.

5.7. Feasibility in Real-Time Gaming Scenarios

The results presented in Table 10 demonstrate the feasibility of deploying our proposed algorithm in real-time gaming scenarios. Across a spectrum of hardware configurations representing diverse computational resources, the algorithm exhibited varying levels of inference speed, latency, and computational demands. In low-end configurations, the algorithm maintained a commendable inference speed of 35 frames per second (fps) with a latency of 60 milliseconds (ms) and GPU usage of 35%, indicating a high responsiveness. As computational resources increased in mid-range and high-end configurations, the algorithm showcased an improved performance, achieving 50 fps with 35 ms latency and 75% GPU usage in high-end setups. These findings suggest that the algorithm adapts well to different hardware conditions, with latency and computational demands aligning with the available resources.

Table 10. Feasibility in real-time gaming scenarios.

Hardware Configuration	Inference Speed (fps)	Latency (ms)	Computational Demands (GPU Usage)	Algorithm Responsiveness
Low-End Configuration	25	60	35%	High
Mid-Range Configuration	50	35	50%	Medium
High-End Configuration	100	20	75%	Low

5.8. Robustness to Noise and Errors Experiment Results

The experimental assessment of the algorithm's robustness to noise and errors within the dynamic gaming environments of Mortal Kombat, Jackie Chan, and Street Fighter reveals promising outcomes, presented in Table 11 below.

Table 11. Robustness to noise and error results.

Game	Task Competition Rate	Policy Stability (Avg)	Adaptability (Success Rate)
Mortal Kombat	92.5%	0.87	85
Jackie Chan	89.2%	0.91	78
Street Fighter	94.8%	0.88	92

The algorithm demonstrated high task completion rates, indicating its effectiveness in achieving game objectives. Policy stability was robust (0.87), ensuring consistent and reliable decision-making. Adaptability, with a 92.5% success rate, highlighted the algorithm's capability to handle dynamic changes in Mortal Kombat's environment. Similarly to Mortal Kombat, the algorithm showcased high task completion rates (89.2%) and strong policy stability (0.91). Despite slightly lower adaptability (78%), the algorithm effectively coped with noise and uncertainties within Jackie Chan's dynamic gaming environment. Street Fighter exhibited exceptional results across all metrics. High task completion rates (94.8%), strong policy stability (0.88), and outstanding adaptability (92%) underscore the algorithm's robustness to noise and errors.

5.9. Overall Implications

The results suggest that the incorporation of the Intrinsic Curiosity Module within the A3C framework provides substantial benefits in enhancing the exploration capabilities of game-playing agents. These findings imply that fostering curiosity-driven exploration enables agents to better understand complex environments, discover optimal strategies, and avoid local optima, ultimately leading to improved performance and adaptability in action games.

The results demonstrated that our curiosity-driven A3C agents consistently outperformed all the baseline approaches in the aforementioned two contexts, namely exploration with limited extrinsic rewards and generalization to novel scenarios. The agents with intrinsic curiosity exhibited an improved exploration efficiency, faster learning speed, and higher average game scores in comparison to vanilla A3C, ICM-pixels, and VIME-TRPO. This confirms the effectiveness of the proposed Intrinsic Curiosity Module in promoting curiosity-driven exploration and enhancing the learning efficiency of game-playing agents in complex action games.

Moreover, the comparison with ICM pixels highlights the significance of learning an embedding that remains unaffected by aspects of the environment that are beyond the agent's control. By utilizing the inverse dynamics model, the curiosity-driven A3C agents learned to extract significant features from the state space. This has resulted in a more efficient exploration and improved adaptation to various scenarios. Overall, the results support the superiority of our curiosity-driven A3C approach in action games, demonstrating its potential as an effective and robust exploration strategy for reinforcement learning agents.

6. Discussion

In this study, we introduced a novel approach for creating curiosity-based intrinsic rewards that adapt well to complex visual inputs. Our approach overcomes the challenge of predicting pixels, which helps keep the agent's exploration strategy strong and unaffected by irrelevant aspects of the environment. By integrating the Intrinsic Curiosity Module (ICM) into the Asynchronous Advantage Actor–Critic (A3C) algorithm, we have showcased substantial enhancements in exploration efficiency and learning speed when compared to the standard methods. By incorporating the Intrinsic Curiosity Module (ICM) into the Asynchronous Advantage Actor–Critic (A3C) algorithm, we have demonstrated significant improvements in exploration efficiency and learning speed compared to baseline approaches.

Our curiosity-driven A3C agent demonstrated outperformance compared to the baseline A3C without curiosity. In addition to a recently proposed exploration formulation known as Variational Information Maximization for Exploration (VIME), there is also a basic formulation that predicts pixels. In the Mortal Kombat, Jackie Chan, and Street Fighter environments, our agent learned to explore and master various combat scenarios without relying solely on external rewards from the game. This showcases the agent's ability to acquire effective exploration behaviors and combat strategies through intrinsic rewards driven by curiosity. While the current research mainly considers single-agent learning, multi-agent reinforcement learning approaches also focus on coordination, collaboration, and competition among multiple agents in complex learning environments. Our approach, in comparison with multiple-agent learning scenarios, may lack in explicit communication or strategy sharing between agents. This may lead to potentially broadening the applicability of our method to collaborative or adversarial environments. Integrating curiosity-driven mechanisms into multi-agent settings remains an open direction for future research.

However, we encountered challenges in certain scenarios, particularly in Street Fighter, where some opponents possessed intricate fighting techniques that required specific sequences of actions to defeat. The agent struggled to receive additional rewards and explore beyond certain limits due to the complexity of these opponent behaviors. This highlights the importance of balancing exploration and exploitation in intricate action games.

To assess the generalization capability of our approach, we evaluated the learned policy on separate 'testing sets' beyond the training environments of Mortal Kombat, Jackie Chan, and Street Fighter. The agent demonstrated promising generalization performance, indicating its ability to transfer learned exploration behaviors to new scenarios and combat environments. This showcases the agent's potential for adaptive learning and its application to novel game environments.

The outcomes attained through 'ICM + A3C' surpassed those of baseline methods ('A3C', 'ICM-pixels + A3C', and 'VIME-TRPO') in terms of mean scores and learning speed. This improvement distinctly signifies the advancement in agents' adaptability and performance within the action games. The study effectively demonstrates that the integration of ICM and A3C techniques contributes significantly to the development of more competent and adaptable game-playing agents.

The study's findings represent a significant leap in reinforcement learning (RL) techniques tailored for action games. By amalgamating curiosity-driven exploration (ICM) with efficient training (A3C), the study introduces a pioneering approach that surpasses traditional RL methods in terms of exploration efficiency, learning speed, and adaptability. These advancements pave the way for the emergence of more intelligent and effective game-playing agents in action-oriented gaming scenarios.

7. Limitations and Challenges

While our study predominantly focuses on the 1v1 game format, for future exploration, a potential avenue involves the algorithm's applicability to multiplayer configurations, such as 2v2 or other team-based setups. The dynamics of multiplayer scenarios introduce additional complexities, including collaborative and competitive interactions among agents. This intriguing dimension offers a rich field for investigation where the algorithm's adaptability to team dynamics, coordination, and strategic decision-making can be scrutinized. Recognizing the potential variations in results and the challenges posed by different multiplayer formats, our study lays the groundwork for future research endeavors to delve into the nuanced dynamics of team-based gaming environments, providing a comprehensive understanding of the algorithm's performance in diverse gaming scenarios.

8. Conclusions

To conclude, it is proven that our curiosity-driven A3C approach has shown promising results in promoting exploration and improving learning efficiency in action games with sparse extrinsic rewards, such as Mortal Kombat, Jackie Chan, and Street Fighter. This happened by converting unexpected interactions and challenges encountered during combat into intrinsic rewards. The agent successfully learned valuable exploration behaviors and combat strategies. While our approach excels in sparse reward settings, there is a potential for further research and extensions to address scenarios with rare opportunities for interactions. For instance, learned behavior can be utilized as a fundamental or low-level policy within a more complex and hierarchical framework, such as a navigation system for exploring intricate game levels and environments. An integration of action–curiosity mechanisms in future work would be a promising direction, as it can enhance the exploration by considering novel states as well as uncertainty and diversity in an agent's actions. In this way, the current trend in reinforcement learning would be aligned with state-of-the-art

methodology and could lead to more efficient exploration strategies. Researchers and practitioners can use our work as a stepping stone to further enhance exploration strategies and reinforcement learning techniques for diverse interactive environments.

Author Contributions: Conceptualization, S.S.F., H.R., S.A.W. (Samiya Abdul Wahid), and H.L.; methodology, S.S.F., M.A.A., S.A.W. (Saira Abdul Wahid), H.R. and H.L.; software, S.S.F. and S.A.W. (Saira Abdul Wahid); validation, S.S.F., S.A.W., H.R. and H.L.; formal analysis, S.S.F. and M.A.A.; investigation, H.R. and H.L.; resources, S.S.F. and H.L.; data curation, S.S.F., S.A.W. (Samiya Abdul Wahid), S.A.W. (Saira Abdul Wahid), and H.L.; writing—original draft preparation, S.S.F., S.A.W. (Samiya Abdul Wahid), S.A.W. (Saira Abdul Wahid), H.R. and H.L.; writing—review and editing, S.S.F., M.A.A., S.A.W. (Samiya Abdul Wahid)., H.R. and H.L.; visualization, S.S.F. and H.R., H.L.; supervision, H.L.; project administration, H.L.; funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Glocal University 30 Project Fund of Gyeongsang National University in 2025 and in part by Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government (MOTIE) 20241K00000010, Robot Utilization Production Support Center For SMR).

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:

ICM	Intrinsic Curiosity Module
A3C	Advantage Actor–Critic (A3C)
FPS	First-Person Shooter
DQNs	Deep Q-Networks
HRL	Hierarchical Reinforcement Learning
LF2	Litter Fighter 2
RL	Reinforcement Learning
MCTS	Monte Carlo Tree Search
ELU	Exponential Linear Unit
LSTM	Long Short-Term Memory
RND	Random Network Distillation
PPO	Proximal Policy Optimization

References

1. Shakya, K.; Pillai, G.; Chakrabarty, S. Reinforcement learning algorithms: A brief survey. *Expert Syst. Appl.* **2023**, *231*, 120495. [CrossRef]
2. Balhara, S.; Gupta, N.; Alkhayyat, A.; Bharti, I.; Malik, R.Q.; Mahmood, S.N.; Abedi, F. A survey on deep reinforcement learning architectures, applications and emerging trends. *IET Commun.* **2022**, *19*, e12447. [CrossRef]
3. Liang, J.; Miao, H.; Li, K.; Tan, J.; Wang, X.; Luo, R.; Jiang, Y. A review of multi-agent reinforcement learning algorithms. *Electronics* **2025**, *14*, 820. [CrossRef]
4. López, K.F.C. Reinforcement Learning Neural Agents in Clever Game Playing. Bachelor's Thesis, Universidad de Investigación de Tecnología Experimental Yachay, San Miguel de Urcuquí, Ecuador, 2022.
5. Han, I.; Kim, K.-J. Deep ensemble learning of tactics to control the main force in a real-time strategy game. *Multimed. Tools Appl.* **2023**, *83*, 12059–12087.
6. Gordillo, C.; Bergdahl, J.; Tollmar, K.; Gisslén, L. Improving playtesting coverage via curiosity driven reinforcement learning agents. In Proceedings of the 2021 IEEE Conference on Games (CoG), Virtual, 17–20 August 2021; pp. 1–8.
7. Chang, O.; Ramos, L.; Morocho-Cayamcela, M.E.; Armas, R.; Zhinin-Vera, L. Continual learning, deep reinforcement learning, and microcircuits: A novel method for clever game playing. *Multimed. Tools Appl.* **2025**, *84*, 1537–1559.

8. Pathak, D.; Agrawal, P.; Efros, A.A.; Darrell, T. Curiosity–driven exploration by self-supervised prediction. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 2778–2787.
9. Mnih, V.; Badia, A.P.; Mirza, M.; Graves, A.; Lillicrap, T.; Harley, T.; Silver, D.; Kavukcuoglu, K. Asynchronous methods for deep reinforcement learning. In Proceedings of the International Conference on Machine Learning, PMLR, New York, NY, USA, 20–22 June 2016; pp. 1928–1937.
10. Mikhaylova, E.; Makarov, I. Curiosity-driven exploration in vizdoom. In Proceedings of the 2022 IEEE 20th Jubilee International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, 15–17 September 2022; pp. 65–70.
11. Knorr, J.W.B.M. Dynamic Difficulty Adjustment in First Person Shooters. Ph.D. Dissertation, Instituto Politecnico do Porto (Portugal), Porto, Portugal, 2021.
12. Chaplot, D.S.; Jiang, H.; Gupta, S.; Gupta, A. Semantic curiosity for active visual learning. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Part VI 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 309–326.
13. Zhang, X. Simulation-based game testing for estimating player curiosity. Master's Thesis, Aalto University, Espoo, Finland, 2023.
14. Yannakakis, G.N.; Hallam, J.; Lund, H.H. Comparative fun analysis in the innovative playware game platform. In Proceedings of the 1st World Conference for Fun'n Games, Preston, UK, 26–28 June 2006.
15. Chen, Y.; Xiao, J. Target search and navigation in heterogeneous robot systems with deep reinforcement learning. *arXiv* **2023**, arXiv:2308.00331. [CrossRef]
16. Sun, C.; Qian, H.; Miao, C. From psychological curiosity to artificial curiosity: Curiosity-driven learning in artificial intelligence tasks. *arXiv* **2022**, arXiv:2201.08300. [CrossRef]
17. Mantiuk, F.; Zhou, H.; Wu, C.M. From curiosity to competence: How world models interact with the dynamics of exploration. 2025. In Proceedings of the 47th Annual Conference of the Cognitive Science Society, San Francisco, CA, USA, 30 July–2 August 2025.
18. Zhong, Y.; He, J.; Kong, L. Double a3c: Deep reinforcement learning on openai gym games. *arXiv* **2023**, arXiv:2303.02271.
19. Ahmad, F.; Shaheen, M.; Ahmed, Z.; Riasat, R.; Muneeb, S. Comprehending the influence of brain games mode over playfulness and playability metrics: A fused exploratory research of players' experience. *Interact. Learn. Environ.* **2023**, *32*, 4881–4897. [CrossRef]
20. Ahmad, F.; Ahmed, Z.; Shaheen, M.; Muneeb, S.; Riasat, R. A pilot study on the evaluation of cognitive abilities' cluster through game-based intelligent technique. *Multimed. Tools Appl.* **2023**, *82*, 41323–41341.
21. Muneeb, S.; Sitbon, L.; Ahmad, F. Opportunities for serious game technologies to engage children with autism in a Pakistani sociocultural and institutional context: An investigation of the design space for serious game technologies to enhance engagement of children with autism and to facilitate external support provided. In Proceedings of the 34th Australian Conference on Human-Computer Interaction, Canberra, ACT, Australia, 29 November–2 December 2022; pp. 338–347.
22. Sulaiman, A.; Rahman, H.; Ali, N.; Shaikh, A.; Akram, M.; Lim, W.H. An augmented reality pqrst based method to improve self-learning skills for preschool autistic children. *Evol. Syst.* **2023**, *14*, 859–872. [CrossRef]
23. Ahmad, F.; Ahmed, Z.; Muneeb, S. Effect of gaming mode upon the players' cognitive performance during brain games play: An exploratory research. *Int. J. Game Based Learn.* **2021**, *11*, 67–76. [CrossRef]
24. Ahmad, F.; Luo, Z.W.; Ahmed, Z.; Muneeb, S. Behavioral profiling: A generationwide study of players' experiences during brain games play. *Interact. Learn. Environ.* **2023**, *31*, 1265–1278. [CrossRef]
25. Tan, W.; Patel, D.; Kozma, R. Strategy and benchmark for converting deep q-networks to event-driven spiking neural networks. In Proceedings of the AAAI conference on artificial intelligence, Vancouver, BC, Canada, 2–9 February 2021; Volume 35, pp. 9816–9824.
26. Girgis, A. Evaluating the efficacy of deep neural networks in reinforcement learning problems. *Am. Sci. Res. J. Eng. Technol. Sci.* **2018**, *46*, 260–272.
27. Lake, B.M.; Ullman, T.D.; Tenenbaum, J.B.; Gershman, S.J. Building machines that learn and think like people. *Behav. Brain Sci.* **2017**, *40*, e253. [CrossRef]
28. Mahmud, A.; Khan, A.K.; Rafi, M.M.H.; Fahim, K.R. Implementation of Reinforcement Learning Architecture to Augment an Ai That Can Self-Learn to Play Video Games. Ph.D. Dissertation, Brac University, Dhaka, Bangladesh, 2023.
29. Nguyen, N.D.; Nguyen, T.T.; Vamplew, P.; Dazeley, R.; Na- havandi, S. A prioritized objective actor-critic method for deep reinforcement learning. *Neural Comput. Appl.* **2021**, *33*, 10335–10349. [CrossRef]
30. Yoon, S.; Kim, K.-J. Deep q networks for visual fighting game ai. In Proceedings of the 2017 IEEE Conference on Computational Intelligence and Games (CIG), New York, NY, USA, 22–25 August 2017; pp. 306–308.
31. Pinto, I.P.; Coutinho, L.R. Hierarchical reinforcement learning with monte carlo tree search in computer fighting game. *IEEE Trans. Games* **2018**, *11*, 290–295. [CrossRef]
32. Takano, Y.; Inoue, H.; Thawonmas, R.; Harada, T. Self-play for training general fighting game ai. In Proceedings of the 2019 Nicograph International (NicoInt), Yangling, China, 5–7 July 2019; p. 120.

33. Takano, Y.; Ito, S.; Harada, T.; Thawonmas, R. Utilizing multiple agents for decision making in a fighting game. In Proceedings of the 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE), IEEE, Nara, Japan, 9–12 October 2018; pp. 594–595.
34. Chakravarthi, B.; Ng, S.-C.; Ezilarasan, M.; Leung, M.-F. Eeg-based emotion recognition using hybrid cnn and lstm classification. *Front. Comput. Neurosci.* **2022**, *16*, 1019776. [CrossRef] [PubMed]
35. Inoue, H.; Takano, Y.; Thawonmas, R.; Harada, T. Verifica- tion of applying curiosity-driven to fighting game ai. In Proceedings of the 2019 Nicograph International (NicoInt), IEEE, Yangling, China, 5–7 July 2019; p. 119.
36. Li, Y.-J.; Chang, H.-Y.; Lin, Y.-J.; Wu, P.-W.; Wang, Y.-C.F. Deep reinforcement learning for playing 2.5 d fighting games. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, Athens, Greece, 7–10 October 2018; pp. 3778–3782.
37. Ishii, R.; Ito, S.; Ishihara, M.; Harada, T.; Thawonmas, R. Monte-carlo tree search implementation of fighting game ais having personas. In Proceedings of the 2018 IEEE Conference on Computational Intelligence and Games (CIG), IEEE, Maastricht, The Netherlands, 14–17 August 2018; pp. 1–8.
38. Bezerra, J.R.; Góes, L.F.W.; Da Silva, A.R. Development of an autonomous agent based on reinforcement learning for a digital fighting game. In Proceedings of the 2020 19th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames), IEEE, Recife, Brazil, 25 December 2020; pp. 1–7.
39. Kim, D.-W.; Park, S.; Yang, S.-I. Mastering fighting game using deep reinforcement learning with self-play. In Proceedings of the 2020 IEEE Conference on Games (CoG), IEEE, Osaka, Japan, 24–27 August 2020; pp. 576–583.
40. Oh, I.; Rho, S.; Moon, S.; Son, S.; Lee, H.; Chung, J. Creating pro-level ai for a real-time fighting game using deep reinforcement learning. *IEEE Trans. Games* **2021**, *14*, 212–220. [CrossRef]
41. Lu, F.; Yamamoto, K.; Nomura, L.H.; Mizuno, S.; Lee, Y.; Thawonmas, R. Fighting game artificial intelligence competition platform. In Proceedings of the 2013 IEEE 2nd Global Conference on Consumer Electronics (GCCE), IEEE, Makuhari, Japan, 4 October 2013; pp. 320–323.
42. Schmidhuber, J. A possibility for implementing curiosity and boredom in model-building neural controllers. In Proceedings of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats; MIT Press/Bradford Books: Cambridge, MA, USA, 1991; pp. 222–227.
43. Schmidhuber, J. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Trans. Auton. Ment. Dev.* **2010**, *2*, 230–247. [CrossRef]
44. Cuccu, G.; Luciw, M.; Schmidhuber, J.; Gomez, F. Intrinsically motivated neuroevolution for vision-based reinforcement learning. In Proceedings of the 2011 IEEE International Conference on Development and Learning (ICDL), IEEE, Frankfurt, Germany, 24–27 August 2011; Volume 2, pp. 1–7.
45. Kearns, M.; Koller, D. Efficient reinforcement learning in factored mdps. In Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 31 July–6 August 1999; Volume 16, pp. 740–747.
46. Brafman, R.I.; Tennenholtz, M. R-max-a general polyno- mial time algorithm for near-optimal reinforcement learning. *J. Mach. Learn. Res.* **2002**, *3*, 213–231.
47. Klyubin, A.S.; Polani, D.; Nehaniv, C.L. Empowerment: A universal agent-centric measure of control. In Proceedings of the 2005 IEEE Congress on Evolutionary Computation, IEEE, Edinburgh, UK, 2–5 September 2005; Volume 1, pp. 128–135.
48. Mohamed, S.; Rezende, D.J. Variational information maximisation for intrinsically motivated reinforcement learning. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 2125–2133.
49. Stadie, B.C.; Levine, S.; Abbeel, P. Incentivizing exploration in reinforcement learning with deep predictive models. *arXiv* **2015**, arXiv:1507.00814. [CrossRef]
50. Osband, I.; Blundell, C.; Pritzel, A.; Van Roy, B. Deep exploration via bootstrapped DQN. In Proceedings of the Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, Barcelona, Spain, 5–11 December 2016; pp. 4026–4034.
51. Osband, I.; Aslanides, J.; Cassirer, A. Randomized prior functions for deep reinforcement learning. In Proceedings of the 32nd International Conference on Neural Information Processing System, Montréal, QC, Canada, 2–8 December 2018; Volume 31, pp. 8626–8638.
52. Mobin, S.A.; Arnemann, J.A.; Sommer, F. Information-based learning by agents in unbounded state spaces. In Proceedings of the 28th International Conference on Neural Information Processing Systems, Montréal, QC, Canada, 8–13 December 2014; Volume 27, pp. 3023–3031.
53. Still, S.; Precup, D. An information-theoretic approach to curiosity-driven reinforcement learning. *Theory Biosci.* **2012**, *131*, 139–148. [CrossRef]
54. Storck, J.; Hochreiter, S.; Schmidhuber, J. Reinforcement driven information acquisition in non-deterministic environments. In Proceedings of the International Conference on Artificial Neural Networks, Paris, France, 27 November–1 December 1995; Volume 2, pp. 159–164.

55. Gao, S.; Ver Steeg, G.; Galstyan, A. Variational information maximization for feature selection. In Proceedings of the Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, Barcelona, Spain, 5–11 December 2016; pp. 487–495.
56. Fang, Z.; Yang, K.; Tao, J.; Lyu, J.; Li, L.; Shen, L.; Li, X. Exploration by Random Distribution Distillation. *arXiv* **2025**, arXiv:2505.11044. [CrossRef]
57. Xue, J.; Chen, J.; Zhang, S. Action-Curiosity-Based Deep Reinforcement Learning Algorithm for Path Planning in a Nondeterministic Environment. *Intell. Comput.* **2025**, *4*, 0140. [CrossRef]
58. Houthooft, R.; Chen, X.; Duan, Y.; Schulman, J.; De Turck, F.; Abbeel, P. Vime: Variational information maximizing exploration. In Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–11 December 2016; Volume 29, pp. 6587–6595.
59. Yang, Z.; Du, H.; Wu, Y.; Jiang, Z.; Qu, H. Intrinsic Motivation Exploration via Self-Supervised Prediction in Reinforcement Learning. In Proceedings of the 6th International Conference on Data-Driven Optimization of Complex Systems (DOCS), Hangzhou, China, 16 August 2024; pp. 79–84.
60. Sun, W.; Cheng, X.; Yu, X.; Xu, H.; Yang, Z.; He, S.; Zhao, J.; Liu, K. Probabilistic Uncertain Reward Model. *arXiv* **2025**, arXiv:2503.22480. [CrossRef]
61. Clevert, D.-A.; Unterthiner, T.; Hochreiter, S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv* **2015**, arXiv:1511.07289.
62. Schmidhuber, J.; Hochreiter, S.; Bengio, Y. Evaluating benchmark problems by random guessing. In *A Field Guide to Dynamical Recurrent Networks*; Wiley-IEEE Press: Hoboken, NJ, USA, 2001; pp. 231–235.
63. Huang, S.; Dossa, R.F.J.; Ye, C.; Braga, J.; Chakraborty, D.; Mehta, K.; Araújo, J.G. Cleanrl: High-quality single-file implementations of deep reinforcement learning algorithms. *J. Mach. Learn. Res.* **2022**, *23*, 12585–12602.
64. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347. [CrossRef]
65. Li, Z.; Ji, Q.; Ling, X.; Liu, Q. A comprehensive review of multi-agent reinforcement learning in video games. *arXiv* **2025**, arXiv:2509.03682. [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.

computers

MDPI

Article

A Guided Self-Study Platform of Integrating Documentation, Code, Visual Output, and Exercise for Flutter Cross-Platform Mobile Programming

Safira Adine Kinari, Nobuo Funabiki *, Soe Thandar Aung and Htoo Htoo Sandi Kyaw

Department of Information and Communication Systems, Okayama University, Okayama 700-8530, Japan; safiraak@s.okayama-u.ac.jp (S.A.K.)
* Correspondence: funabiki@okayama-u.ac.jp

Abstract: Nowadays, *Flutter* with the *Dart* programming language has become widely popular in mobile developments, allowing developers to build multi-platform applications using one codebase. An increasing number of companies are adopting these technologies to create scalable and maintainable mobile applications. Despite this increasing relevance, university curricula often lack structured resources for Flutter/Dart, limiting opportunities for students to learn it in academic environments. To address this gap, we previously developed the *Flutter Programming Learning Assistance System (FPLAS)*, which supports self-learning through interactive problems focused on code comprehension through code-based exercises and visual interfaces. However, it was observed that many students completed the exercises without fully understanding even basic concepts, if they already had some knowledge of *object-oriented programming (OOP)*. As a result, they may not be able to design and implement *Flutter/Dart* codes independently, highlighting a mismatch between the system's outcomes and intended learning goals. In this paper, we propose a *guided self-study approach of integrating documentation, code, visual output, and exercise* in *FPLAS*. Two existing problem types, namely, *Grammar Understanding Problems (GUP)* and *Element Fill-in-Blank Problems (EFP)*, are combined together with documentation, code, and output into a new format called *Integrated Introductory Problems (INTs)*. For evaluations, we generated 16 *INT* instances and conducted two rounds of evaluations. The first round with 23 master students in Okayama University, Japan, showed high correct answer rates but low usability ratings. After revising the documentation and the system design, the second round with 25 fourth-year undergraduate students in the same university demonstrated high usability and consistent performances, which confirms the effectiveness of the proposal.

Keywords: Flutter; Dart; cross-platform; self-learning; introductory

1. Introduction

Nowadays, various devices such as personal computers, smartphones, tablets, smart watches, and smart eyeglasses are used with different software platforms and operating systems including *Windows*, *Linux*, *Android*, and *iOS* [1,2]. As a result, the strong demand for cross-platform developments has been growing in IT industries around the world [3]. Traditionally, building applications for different platforms required separate implementations using native programming languages, since each platform has its own structure and features [2]. For example, *Android* development typically uses *Java* or *Kotlin*, *iOS* uses *Swift* or *Objective-C*, and web applications are developed using *HTML* and *JavaScript*.

Under such situations, a service vendor needs to develop and maintain separate codebases using different programming languages for each target platform [4]. This typically results in higher development costs, broader team skill requirements, and increased maintenance overhead [5].

To solve this undesired situation, developers around the world have started using cross-platform development frameworks that allow a single codebase to run on different devices/platforms [6]. On such a framework, a service vender needs to implement and update one code using one language for one service. Then, the framework outputs the runnable code on each specific device/platform from it, which makes application developments in heterogenous environments much easier. One of the most popular frameworks is *Flutter* [7].

Developed by *Google*, *Flutter* can provide a solution in fast cross-platform application developments without sacrificing system performances or visual quality. *Flutter* is powered by the *Dart* programming language [8], which is also developed by *Google*, and offers several developer-friendly features, such as a unified codebase for *Android*, *iOS*, *web*, and *desktop* applications, a flexible *widget* system, and a *hot reload mechanism* for real-time iteration [9,10].

However, *Flutter* has been limited in academic settings, although it provides many benefits for developers. Most university curricula in the world have not yet included *Flutter* and *Dart* as part of their programming courses. As a result, students are facing challenges in gaining practical experience using *Flutter* for application developments. To address this gap, self-learning tools will be efficient solutions that offer proper guidance and hands-on practice, enabling students to study *Flutter/Dart* effectively on their own.

Aiming to support self-directed learning, we have developed the *Programming Learning Assistant System (PLAS)* as a web-based exercise platform for independent programming study. *PLAS* provides various types of exercise problems to help students build both conceptual understanding and practical skills. Then, we have extended *PLAS* to create the *Flutter Programming Learning Assistant System (FPLAS)* [11], which focuses on helping students practice *Flutter/Dart*-based application development.

FPLAS integrates multiple learning components. This environment is set up using *Docker container platform* to ensure consistent executions across systems [12]. In *FPLAS*, the *Grammar-Concept Understanding Problem (GUP)* focuses on syntax and grammar fundamentals in *Dart* [11]. Meanwhile, the *Code Modification Problem (CMP)* allows students to practice editing existing *Flutter* code [13].

During the initial use of *FPLAS*, unfortunately, we observed an unexpected pattern among the participants of the evaluations. Certain students were able to answer questions correctly, despite lacking a full understanding of fundamental *Flutter/Dart* concepts. This was particularly common among those with prior experiences in *JavaScript*, *Java*, or *object-oriented programming (OOP)*.

To provide structured supports, we developed beginner-oriented documentation that introduces essential concepts using simple explanations and example codes to help students understand the fundamentals of *Flutter/Dart* [14]. Then, based on this documentation, we designed a new set of problem instances called the *Integrated Introductory Problem (INT)*, and implemented them in *FPLAS*. The *INT* problems adopt selected formats from *PLAS*, such as the *Grammar Understanding Problem (GUP)* and the *Element Fill-in-blank Problem (EFP)*. In addition, the *INT* provides interface screenshots to help students relate the code to the interface layout. the *INT* aims to strengthen students' conceptual understanding by linking learning contents with hands-on coding practices, thereby helping bridge the gap between theory and implementation.

An *INT* instance combines documentation, code editing, visual output, and interactive exercise problems into a single platform. This setup enables students to read explanations, modify the code, and see the visual screen results. The *INT* helps them connect abstract concepts with practical implementations, which is especially important in User Interface (UI)-focused frameworks, including *Flutter*.

To evaluate the proposal, we generated *INT* instances to study key concepts in the developed documentation, and conducted two rounds of user testing. In the first round, we prepared 16 instances, and assigned them to 23 master's students at Okayama University, Japan, who had no prior experience with *Flutter* or *Dart*. The purpose of this round was to assess the system's usability and educational effectiveness. These students were generally able to complete the assignments correctly, indicating a basic understanding of the targeted programming concepts. However, the *System Usability Scale (SUS)* results suggested that the usability of the system could be improved [15].

Based on the feedback in the first round, we revised both the documentation and the system interface. Improvements were made to the clarity of explanations, navigation structure, and problem instructions to make the learning process smoother and more intuitive for beginners.

In the second round, the revised version of the system was evaluated with a different group of 25 fourth-year undergraduate students. The *SUS* scores were improved, showing better user satisfactions, and the solutions remained consistently accurate. Additional feedback in this round highlighted further areas for improvements, which we used to make additional refinements to the learning environment and instructional flow.

To systematically evaluate the proposal, we address the following research questions.

- **RQ1:** Does the integration of structured documentation, code, visual output, and *INT* problems improve students' understanding of *Flutter/Dart* concepts?
- **RQ2:** How does the revised system affect the system's usability and user satisfaction, as measured by the *System Usability Scale (SUS)*?

The structure of the paper is outlined as follows: Section 2 contains a review of relevant research on the approach. Section 3 introduces the beginner-oriented documentation. Section 4 describes the integrated learning design and the rationale for the *Integrated Introductory Problem (INT)*. Section 5 presents the design and implementation of the proposed learning assistant system. Section 6 presents the experimental results. Section 7 discusses student feedback and system improvements. Section 8 discusses the findings, limitations, and implications. Section 9 concludes this paper with future work.

2. Related Works

In this section, we review existing literature associated with our research topic.

2.1. Flutter for Learning

Several studies have explored the use of *Flutter* in educational contexts, highlighting its technical capabilities, architectural clarity, and suitability for cross-platform developments.

Nagaraj et al. [16] emphasized *Flutter*'s widget system and hot reload functionality, which support fast iteration and interactive learning. Zou et al. [17] focused on developer experiences, noting that asynchronous behaviors, animation memory managements, and network image handling often confuse novice learners. These studies point to the need for guided scaffolding when introducing *Flutter* to beginners.

Boukhary et al. [18] proposed a clean architecture model in *Flutter* to separate UI components from business logic, aiming for maintainable and scalable educational projects. Radadiya et al. [19] further elaborated on *Flutter*'s rendering pipeline, widget layering, and integration with IDEs, which are relevant when teaching structured UI developments. The

architectural overview in [20] offers a systematic explanation of *Flutter*'s framework and engine layers that can serve as instructional material for foundational learning.

Sharma et al. [21] demonstrated how *Flutter* enables students to build mobile applications for multiple platforms without maintaining separate codebases. This efficiency makes it an attractive option for introductory courses with limited time or resources.

While these works demonstrate *Flutter*'s educational potential, most of them focus on developer experiences or advanced use cases. They often assume prior programming knowledge and do not address the conceptual difficulties faced by beginners in transitioning from imperative programming styles to declarative UI paradigms.

In contrast, our proposed system focuses on supporting beginners by emphasizing fundamental concepts such as widget hierarchy, *Dart* syntax, reusable widget components, and code structure patterns. The learning materials are carefully aligned with these core elements to provide a structured and progressive learning path that mirrors how *Flutter* is introduced in academic settings.

2.2. Programming Educational Platforms

In addition to studies that focus on the educational use of *Flutter*, various programming learning platforms have been developed to support structured instructions, automated code assessments, and learner feedback across different contexts.

Wang et al. [22] introduced *CodingHere*, a browser-based programming system that emphasizes automatic code assessments with error-specific feedback. Their platform enhances student engagements by providing flexible code checking and diagnostic messages tailored to common mistakes.

Hinrichs et al. [23] developed *OPPSEE*, an online platform equipped with full IDE support and automated feedback. Built on a microservice architecture and integrated with *Gitpod*, it supports multi-file assignments and collaborative programming in real time, offering flexibility for instructors and learners.

To better understand the landscape of platforms, Combéfis et al. [24] conducted a broad classification and comparison of automated code assessment tools. Their analysis identified dimensions such as feedback granularity, language coverages, and analysis methods, providing a taxonomy useful for educational tool designers.

Other studies have emphasized conceptual understanding. Jevtić et al. [25] employed *Self-Organizing Maps (SOMs)* to analyze the density and distribution of programming concepts in instructional material. This method offers visualization-based feedback on curriculum design and concept progression.

A different approach was proposed by Hsueh et al. [26], who developed *Pytutor*, a platform that evaluates both student codes and the quality of generated test cases. Their results suggest that students with stronger testing habits achieve better conceptual outcomes, highlighting the value of integrating test-driven learning.

Emerging directions explore accessibility and mobile learning. Hasan et al. [27] introduced a mobile-focused evaluation platform that supports app developments in *Java* and *Kotlin*. Their system uses test case matching and automated result comparisons to offer real-time feedback on student submissions, aligning well with mobile development scenarios.

Togni et al. [28] proposed an early-stage system that integrates real-time speech recognition, *YOLOv5*-based object detection, and a cross-platform *Flutter* interface. Although this work is currently available as a preprint and under peer review, it reflects the growing interest in adaptive, Artificial Intelligence (AI)-driven educational tools for inclusive mobile learning environments.

While these systems offer a variety of functionalities, from automatic grading and feedback to visual analytics and mobile support, many of them focus on general-purpose

programming or system scalability. Their instructional designs, although effective, are not always tailored to the specific challenges of UI-focused, declarative frameworks such as *Flutter*.

In contrast to these systems, the proposed *INT* format integrates documentation, code, output images, and interactive exercises into a unified instructional instance. This design draws inspirations from pedagogical principles of worked examples [29], while extending them through syntax-level scaffolding (*GUP*) and immediate applications via fill-in-the-blank programming tasks (*EFP*).

Unlike conventional worked examples, which often present complete solutions for passive observations, the *INT* format promotes active engagements and progressive concept acquisitions by guiding learners from recognition to production. It embodies principles of scaffolded learning and just-in-time instruction, helping students build both conceptual and procedural knowledge with reduced cognitive load.

3. Overview of Flutter Introductory Documentation

In this section, we introduce the beginner-oriented documentation for *Flutter/Dart*.

3.1. Purpose and Development

To support novice learners in understanding *Flutter/Dart* programming concepts, we developed an introductory documentation tailored to beginners. This documentation was created in response to the frequent difficulties observed among students during their initial exposures to *Flutter/Dart*. Many students tend to carry over object-oriented and imperative patterns from languages such as *Java* or *JavaScript*, which leads to misunderstandings of *Flutter*'s declarative UI structure.

For example, students often do not understand why they need to define a `main()` function or why `runApp()` must be called before anything is displayed. They may expect the visual interface to appear automatically or try to update elements directly, rather than understanding that UI components are rebuilt via the `build()` method, which must remain free of side effects. Such misconceptions highlight the need for clearer explanations of Flutter's rendering flow, widget hierarchy, and state management concepts.

The aim of this documentation is to reduce this gap by providing learning materials that use simpler language, visual supports, and targeted explanations of fundamental components.

The documentation is designed not only to introduce *Flutter/Dart* syntax and components but also to offer a structured path that gradually builds student understanding, starting from the project structure and moving toward dynamic widget interactions and state managements.

3.2. Documentation Content and Its Role in INT

The documentation consists of 11 core topics that are selected to guide beginners through both *Dart* language foundations and *Flutter* application structures. These topics are adapted from our previous work [13] and are restructured to support this learning system. These topics include the following:

- Basic *Dart* programming concepts such as variables, data types, operators, null safety, generics, error handling, and *object-oriented programming (OOP)*;
- The role of widgets as the core unit in *Flutter*'s architecture;
- Using *constructor* arguments in the key-value format to configure widgets;
- Composition using nested widgets and layout nesting;
- Treating functions and properties as values within *Flutter*;
- Overview of the *Flutter* and *Dart* code structure;

- Recognizing various elements that can be used as values;
- Conditional rendering using ternary operators;
- Understanding the hierarchy of widgets in a *Flutter* interface;
- The difference between *StatelessWidget* and *StatefulWidget*;
- Updating the interface dynamically using the `setState()` method.

Each topic is accompanied by an annotated source code and rendered output screenshots to facilitate intuitive understanding. The full structure and content of this documentation have been presented in our previous work [14], as a self-study material for *Flutter* beginners.

In this study, the documentation serves as the foundation for designing the *Integrated Introductory Problem (INT)*. Topics from the documentation are mapped to individual *INT* instances, each integrating explanation, a sample code, and interactive exercises. This mapping ensures that the learning flow remains consistent and that exercises closely align with the conceptual structure introduced in the documentation.

4. Integrated Learning Design for Flutter Self-Study

In this section, we present a new learning format called the *Integrated Introductory Problem (INT)*. This format builds on the documentation introduced in Section 3, and combines conceptual guidance, a source code, an interface visual output, and interactive exercises into a unified format. The goal of *INT* is to provide a structured way for students to move from reading to practice while reinforcing key programming principles.

4.1. Rationale for Improvement

Our previous *FPLAS* includes the *Code Modification Problem (CMP)* and the *Code Writing Problem (CWP)*. These formats provided mobile application screenshots and problem instructions that will guide students to complete given programming challenges.

Although many students successfully solved the problems, we observed that students with prior experiences in *OOP* languages such as *JavaScript* or *Java* solved them by relying on coding habits. This means that this previous approach in *FPLAS* might lead them to answer questions without fully understanding how *Flutter/Dart* structures the programs and manages the user interfaces. As a result, they often showed limited grasp of the underlying concepts and faced difficulties when tackling more complex and practical programming exercises.

4.2. Overview of INT Framework

To better connect conceptual learning with hands-on coding exercises, in this paper, we design the *Integrated Introductory Problem (INT)* format. Each *INT* instance integrates the following four elements:

- Explanations or summaries that are beginner-friendly based on the guide documentation;
- Source code that directly reflects the explained concept;
- A screenshot of the output screen that visually represents the expected app output;
- Two interactive task types:
  - *Grammar Understanding Problem (GUP)*—questions designed to test understanding of structure, logic, and key terms in *Flutter/Dart*;
  - *Element Fill-in-Blank Problem (EFP)*—partial code with blanks that students must complete to match the expected codes.

This *INT* format connects written explanations on code behaviors and visual outcomes, encouraging students to understand both how the code works and why it behaves that way. While *GUP* and *EFP* offer valuable hands-on experiences, the *INT* format is designed

to further support conceptual understanding by more explicitly linking documentation contents, source codes, and the resulting interfaces. Rather than focusing only on implementations, this alignment helps learners form a clearer connection between what they read, what they write, and what they observe on a screen.

4.3. Curriculum Design and Topic Selection

The 16 topics listed in Table 1 were derived from the 11 foundational topics that are covered in the beginner-oriented documentation introduced in Section 3. These topics were based on difficulties observed in previous *FPLAS* exercises, such as *Code Modification Problems (CMPs)* and *Code Writing Problems (CWPs)*, where students struggled to understand *Flutter*'s declarative structure and widget hierarchy. To address these issues, the 11 fundamental topics were expanded and reorganized into 16 fine sections. This restructuring enables a step-by-step learning experience, helping students build conceptual understanding before moving on to more advanced topics.

Table 1. Topics of *INT* instances.

ID	Topic	Key Concept in Focus	Related Documentation Topic	Purpose in INT System
1	Import and Main	Application entry point using `main()`	Basic *Dart*/Project Structure	Introduce where and how *Flutter* apps start
2	Structure of Project	File and class organization	Project Structure	Help students navigate and understand basic file layout
3	MaterialApp and Text	Root widget setup and rendering simple text	Widget Usage/Project Structure	Demonstrate how a minimal screen is constructed
4	Basic Key-Value	Passing parameters through constructors	Key-Value Format	Emphasize configuration through parameterized widgets
5	Scaffold and Hierarchy	Layout and nesting via `Scaffold`	Widget Hierarchy	Clarify structural containers and layout relationships
6	Change Text Style	Customizing visual output via widget properties	Using Properties as Values	Show how to change the appearance of basic widgets
7	Using AppBar	Adding common UI components	*Flutter* Widget Usage	Introduce real-world components in UI design
8	Using Container	Layout control and decoration	Layout and Nesting/ Container Use	Demonstrate flexible styling and positioning
9	Row and Column	Horizontal and vertical alignment of widgets	Widget Hierarchy	Teach composition and alignment principles
10	Button	Introducing interactivity	Stateless/StatefulWidget	Begin adding user interaction elements
11	*Dart* List	Generating UI elements from lists	*Dart* List	Reinforce repeated UI from data structures
12	Basic Function	Function declaration and reuse	*Dart* Functions	Practice organizing logic outside widget trees
13	StatefulWidget	Handling dynamic values in UI	StatefulWidget Usage	Lay foundation for apps with changing states
14	Button Click Function	Executing logic on interaction	Event Handling	Connect UI actions to *Dart* functions

Table 1. *Cont.*

ID	Topic	Key Concept in Focus	Related Documentation Topic	Purpose in INT System
15	Update with setState	Updating interface with internal state change	setState Mechanism	Show how state affects visible UI
16	Ternary Operator	Conditional UI rendering in expressions	Conditional Expression	Introduce short syntax for visual condition control

Each topic was then implemented as an *INT* instance that integrates the documentation, visual outputs, and interactive exercises. This sequence starts with basic structural concepts of a *Flutter/Dart* application and progressively moves to more advanced UI behaviors and state managements. Table 1 summarizes the topics, the key concepts, and their pedagogical role in the system.

The order of these topics was carefully designed to reflect the dependency of the concepts. Topics like state changes and widget rebuilding are placed later in the sequence, after learners have gained familiarity with widget structure and composition. This careful sequencing helps maintain the logical flow of learning and prevents students from being overwhelmed by advanced behaviors before mastering the foundational syntax and patterns.

By combining structured guidance, example codes, visual results, and targeted questions, the *INT* format enables students to follow a consistent learning iteration that reinforces understanding through repeated application. This integrated design offers a more effective alternative to traditional code-only programming tasks.

5. Integrated Introductory Problem for Flutter/Dart Programming

In this section, we present how the *Integrated Introductory Problem (INT)* is implemented within the *PLAS* for self-learning of *Flutter/Dart*. Each *INT* instance is designed to combine conceptual explanation, code analysis, and structured practice problems in a cohesive and accessible format.

5.1. System Architecture and Workflow

Each *INT* instance integrates two types of exercise problems in *PLAS*: the *Grammar Understanding Problem (GUP)* and the *Element Fill-in-Blank Problem (EFP)*. The *GUP* focuses on conceptual recognitions by asking learners to identify key elements such as keywords, widget names, or method names in given code samples. The *EFP*, on the other hand, requires students to complete partially hidden code based on both the surrounding syntax and the intended output. To solve an *INT* problem, students are presented with the following:

- A partial source code file embedded with blanks;
- Supporting explanations from the documentation (Section 3);
- Visual output in the form of screenshots showing the intended behavior of the app.

This workflow combines different forms of learning, such as reading explanations, writing codes, and observing their application interfaces, into a single interactive process. Although students may rely on trials and errors at times, the system encourages them to carefully examine the code and refer to the documentation when necessary.

5.2. Procedure for Creating INT *Instances*

Each *INT* instance can be generated through the following structured authoring process:

1. A complete *Flutter/Dart* source code is first prepared, targeting one of the selected instructional topics.
2. The source code is executed, and the screenshots of the rendered mobile application interface are captured.
3. Specific keywords, expressions, or syntax elements are manually selected to be blanked out for the *GUP* and *EFP* formats.
4. These blanks and their positions are recorded in the configuration file, which is then processed by a custom *Python* script to generate the problem set.
5. The script produces a self-contained interface using *HTML*, *CSS*, and *JavaScript*, embedding the source code with blanks, input fields, and screenshots in a layout suitable for web-based use.

Since the current blank selection process is manual, this step offers flexibility in targeting specific learning objectives. In each *INT* instance, the blanked elements in the *EFP* were chosen to match a keyword, a method, or a grammar element introduced in the corresponding *Grammar Understanding Problem (GUP)* that precedes it. Future works will explore the automation of this step using rule-based or context-aware blanking strategies.

5.3. Example Instance: Topic #2 – Flutter Code Structure

An example instance is shown in Listing 1, where students are introduced to the basic structure of a *Flutter/Dart* application code. This code demonstrates the use of a *stateless widget* class and how the application starts using the `main()` function. Its related *GUP* questions ask students to identify class names, *override* annotations, and lifecycle methods.

Listing 1. Source code for INT instance on project structure.

```
import 'package:flutter/material.dart';

void main() {
  runApp(MyApp());
}

class MyApp extends StatelessWidget {
  MyApp({super.key});

  @override
  Widget build(BuildContext context) {
    return FlutterLogo();
  }
}
```

Table 2 shows sample *GUP* questions derived from this code. These questions were manually authored by the researchers to reflect common keyword-level misunderstandings encountered in early-stage Flutter learning. No automatic question generation was used in this version.

The corresponding *EFP* input file (Listing 2) replaces key elements with double-dollar markers such as `$main$`. These are parsed into input fields.

Table 2. Sample *GUP* questions and answers.

No.	Question	Answer
1	Which keyword represents the main application widget class?	`MyApp`
2	Which keyword is used to override a method in *Flutter*?	`@override`
3	Which keyword is used to send a value back from a function?	`return`

Table 2. *Cont.*

No.	Question	Answer
4	Which method is used to define the UI layout of a widget?	`build`
5	Which keyword allows a class to inherit from another?	`extends`
6	Which syntax is used to initialize a superclass's key?	`super.key`

Listing 2. Input code with blanks for EFP format.

```
[QUESTION]
import '$package$:flutter/material.dart';

void $main$() {
  runApp(MyApp());
}

class MyApp extends StatelessWidget {
  MyApp({super.key});

  @override
  Widget build(BuildContext context) {
    return FlutterLogo();
  }
}
```

5.4. Interface Design and Scoring Mechanism

As shown in Figure 1, the *INT* web answer interface is divided into two panels: the left panel displays the source code with input fields, and the right panel shows the expected visual output. Students enter their answers directly into the blanks and click a button to receive feedback.

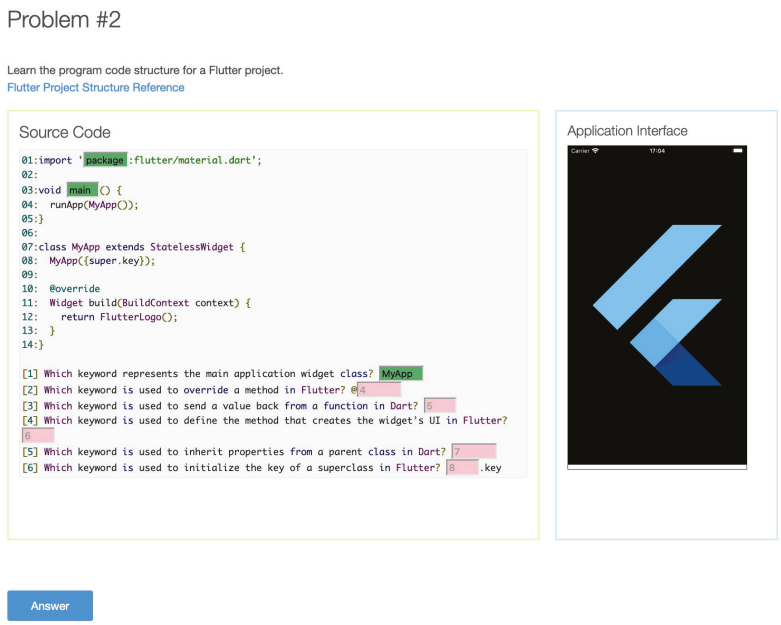


Figure 1. Answer interface for *INT* instance.

The validation is performed by a client-side JavaScript function that checks each input against the correct answer using the exact *string matching*. If the answer is correct, the input

field turns green. The incorrect answer turns it red. Students may revise and resubmit their answers as many times as needed.

To prevent access to correct answers, all answers are hashed using *SHA256*, not stored in a plain text. This enables both online and offline use while preserving fairness in open-access environments.

6. Evaluation

In this section, we conduct an evaluation of the proposal to examine its effectiveness by generating 16 *INT* instances covering fundamental *Flutter/Dart programming* topics and giving them to students at Okayama University, Japan.

6.1. Evaluation Setting

To evaluate the effectiveness and usability of the proposed *FPLAS* system with *Integrated Introductory Problems (INTs)*, we conducted two rounds of evaluations involving student participants from Okayama University, Japan. Each round examined answer correctness to assigned problems and usability impressions to assess how the system supports conceptual learning and user experiences.

6.1.1. First Round Participants (Initial Group)

The first evaluation was conducted with 23 master's course students in the *Department of Information and Communication Systems* at Okayama University. Most of them had completed their undergraduate studies in the same department, ensuring a consistent academic background. All the participants had previously completed an introductory programming and a follow-up *object-oriented programming (OOP)* course during bachelor's studies, typically using languages such as *Java 21*, *C++20*, or *Python 3.12*. These students participated in the evaluation by solving the *INT* problems using the initial version of the system before any interface or documentation improvements were applied.

6.1.2. Second Round Participants (After-Modification Group)

The second evaluation involved 25 fourth-year undergraduate students (final-year bachelor's students) in the same department at Okayama University. These students had completed both foundational programming and a follow-up *OOP* course as part of their regular curriculum, using the same set of programming languages as the first group. Although they were in the undergraduate stage, their programming educations were broadly comparable to those of the master's students.

The similarity in institutional context, department, and curriculum structure ensures that both groups share a comparable academic foundation, making them appropriate for use in consecutive evaluation rounds. While the two groups differ in the academic seniority, neither group had received advanced programming instruction during the coursework, allowing us to reasonably treat them as similar subsets for the purpose of this exploratory system evaluation. These participant groups were selected based on the available formal class sessions in which the system could be tested.

6.1.3. Evaluation Procedure

To conduct the evaluation, we assigned the proposal to students as a two week task. Students accessed the system through a web browser on their personal devices, without any restriction on the platform or operating system. No *Docker* environment was required, as the system was hosted on a standard server and delivered via a browser interface.

Before the assignment, students attended a 2 × 50-min class session where we introduced the documentation, explained the *INT* problem format, and demonstrated a use of

the system. This ensured that participants had sufficient guidance prior to engaging with the exercises independently.

Students were free to use external resources, including the documentation and Internet searches, as the goal was to replicate realistic self-study conditions. All exercise answers and usage logs were recorded for analysis.

6.2. Solution Results

Here, we discuss the solution results by the students.

6.2.1. Results of Individual Students

First, we discuss the solution results for the individual students. In this context, *student submissions* refers to each student's total number of attempts submitted to solve an *INT* instance, including incorrect ones. The *number of answer submission times* thus represents how many times a student interacted with the system before achieving the correct answer. Figure 2 shows the correct answer rate and the number of answer submission times for each of the 23 master's students among the 16 *INT* instances. This table shows that 17 students among 23 achieved a perfect score (100%), and 6 students achieved a score above 89%. The average number of student submissions per instance was 5.8. For the instances in which students obtained a perfect score (100%), the average number of submissions was 6.7. The results indicate that the generated *INT* instances effectively served as learning tools for *Flutter/Dart* for novice students.

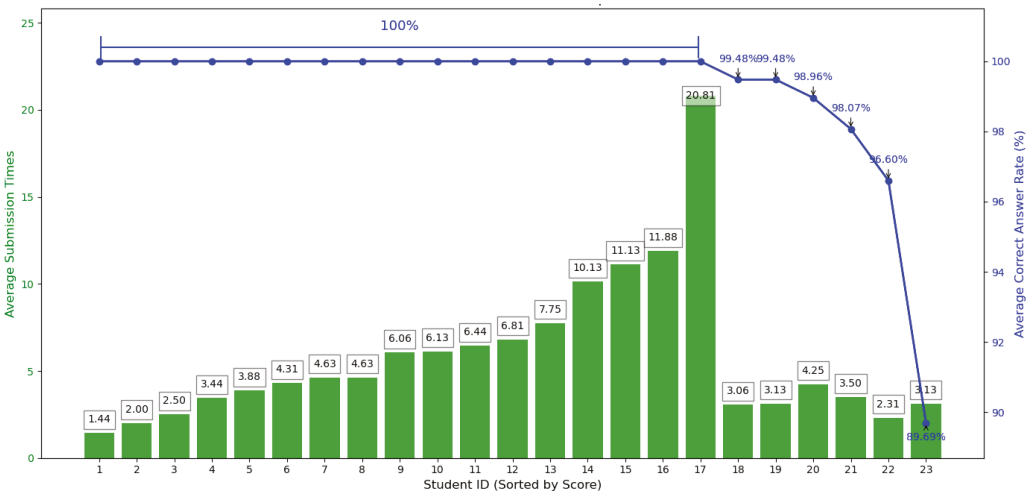


Figure 2. Average correct answer rate and number of submission times for each student's Flutter INT.

6.2.2. Results of Individual Instances

Then, we discuss the solution results for the individual instances. Figure 3 shows the average correct answer rate and the average number of answer submission times among the 23 master's students for each of the 16 *INT* instances. The result suggests that the difficulty varied across the instances, where the lowest rate was 97.10% for ID = 13 and the highest rate was 100% for multiple IDs such as 1, 2, 3, 6, 7, 14, and 15. Similarly, the number of answer submission times required also varied, where the lowest one was 3.13 for ID = 3 and the highest one was 15.04 for ID = 13. Instance ID = 13 exhibited both high submissions and low correct rate due to students' difficulties in understanding state and build concepts.

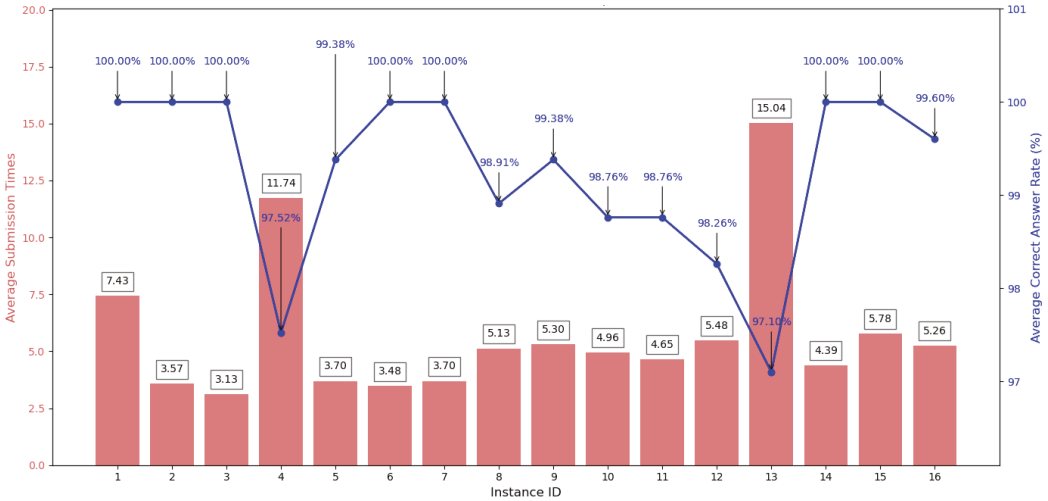


Figure 3. Average solution results for Flutter INT.

6.3. Solution Results After Modification

Some critical feedback was obtained, which was also indicated by the *SUS* scores presented in the next subsection. In response, we made several minor refinements to the instance interface. These included adding location-specific guidance within the documentation, incorporating Japanese translations alongside the original English content, and rephrasing a few unclear questions while keeping the total question count unchanged. Subsequently, we conducted a second evaluation of the same 16 *INT* instances with 25 students from the same university.

6.3.1. Results of Individual Students

We now present the results from the second evaluation. Figure 4 illustrates the average correct answer rate and the average number of answer submissions for each of the 25 fourth-year undergraduate students who attempted the 16 *INT* instances. According to the data, 21 out of 25 students achieved a perfect score (100%), while the remaining 4 students scored above 93%. On average, each student submitted answers 4.4 times per instance. Among those who achieved full marks, the average number of submissions was 4.52. These findings suggest that the revised *INT* instances continued to function effectively as learning materials.

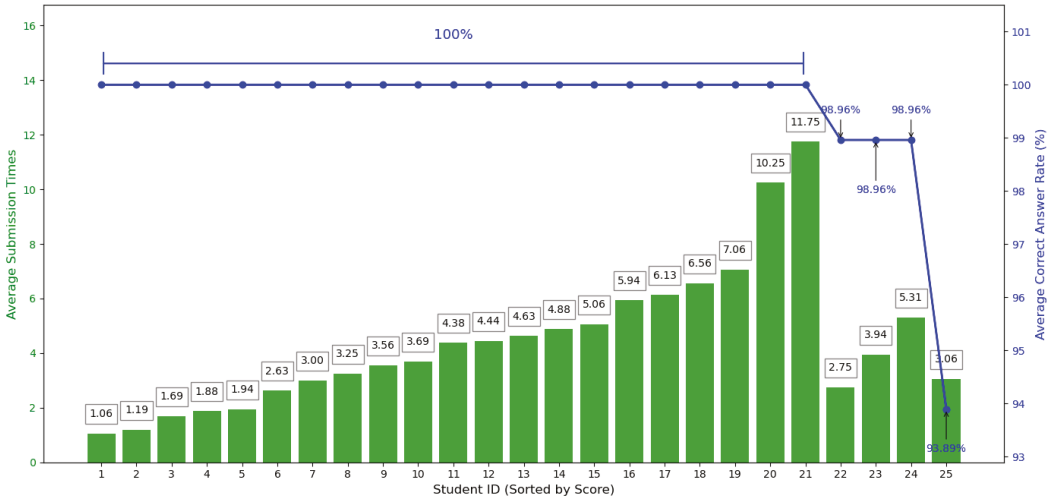


Figure 4. Average correct answer rate and number of submission times for each student's Flutter INT after modification.

6.3.2. Results of Individual Instances

Next, we present the solution results for the individual instances in the second evaluation. Figure 5 shows the average correct answer rate and the average number of answer submission times among the 25 fourth-year undergraduate students for each of the 16 *INT* instances. This suggests that the difficulty still varied across the instances, where the lowest rate was 97.33% for ID = 13 and the highest rate was 100% for multiple IDs such as 3, 4, 5, 7, 9, 10, 11, 12, and 14. Similarly, the number of answer submission times also varied, where the lowest one was 2.36 for ID = 14 and the highest one was 13.96 for ID = 13.

Compared to the first evaluation, the results demonstrate an overall improvement in both the correct answer rate and the number of answer submissions. The average correct answer rate increased from 99.23% to 99.63%, while the average number of submissions per instance decreased from 5.80 to 4.40. These trends suggest that the minor revisions had a positive impact on student performance and understanding. Although instance ID = 13 remained the most challenging, with the highest number of submissions and the lowest accuracy in both evaluations, a slight improvement was still observed, with a $+0.23\%$ increase in the correct answer rate and a -1.08 reduction in the average number of submissions. In contrast, instance ID = 4 showed significant progress, reducing its average submissions from 11.74 to 2.96 and achieving a perfect correct answer rate. This notable improvement indicates that the applied revisions were particularly effective in clarifying the content of certain topics.

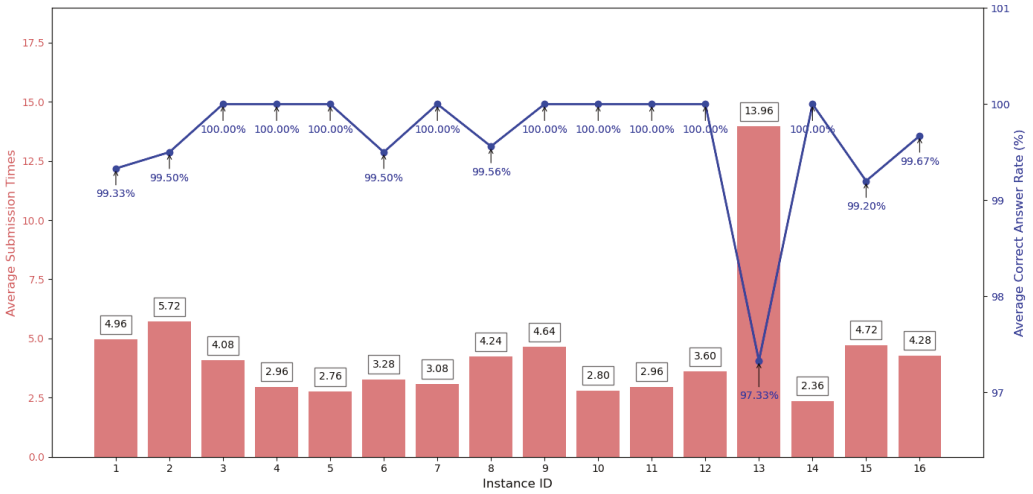


Figure 5. Average solution result for Flutter INT after modification.

6.4. System Usability Scale

To evaluate the usability of the proposal, we used the *System Usability Scale (SUS)* [15], a widely used tool consisting of 10 standardized questions rated on a 5-point Likert scale. This instrument captures aspects such as ease of use, consistency, and learnability.

6.4.1. SUS Question Design

Table 3 lists the *SUS* questions adapted for this study. The questions consist of both positive and negative statements related to the system usability, including flow, clarity of documentation, and helpfulness of feedback.

The scoring procedure follows the standard *System Usability Scale (SUS)* method that was introduced by Brooke [15]. Questions Q1, Q3, Q5, Q7, and Q9 are positively worded

and scored as (response − 1), while Q2, Q4, Q6, Q8, and Q10 are negatively worded and reverse-scored as (5 − response). The final *SUS* score is calculated as:

$$SUS_score = \left(\sum_{i=1}^{5} (Q_{odd_i} - 1) + \sum_{j=1}^{5} (5 - Q_{even_j}) \right) \times 2.5$$

Table 3. SUS questions in evaluation.

ID	Question Statement
Q1	The system helped me clearly understand *Flutter* concepts through its exercises and examples.
Q2	I found the system overly complex to navigate.
Q3	I thought the system was intuitive and easy to use for learning *Flutter*.
Q4	I frequently found it difficult to follow the system's learning flow.
Q5	The system's learning flow (e.g., tutorials, exercises, feedback) felt logical and easy to follow.
Q6	I felt there were inconsistencies in the system's design or functionality.
Q7	The accompanying documentation provided a clear and helpful introduction to *Flutter* concepts.
Q8	I felt the documentation was confusing or lacked clarity in presenting *Flutter* concepts.
Q9	The documentation explained *Flutter*'s core principles in a way that was easy to understand.
Q10	I thought the explanations of *Flutter*'s core principles were overly technical or hard to grasp.

6.4.2. SUS Score Interpretation

SUS scores range from 0 to 100, and can be mapped to letter grades that correspond to the perceived usability level. Table 4 summarizes the grade mapping used in this evaluation.

Table 4. SUS score to grade mapping.

SUS Score Range	Grade
$85 \leq x \leq 100$	A (Best Imaginable)
$80 \leq x < 85$	B (Excellent)
$70 \leq x < 80$	C (Good)
$50 \leq x < 70$	D (OK)
$25 \leq x < 50$	E (Poor)
$x < 25$	F (Worst Imaginable)

6.4.3. Usability Result Before Modification

Before we applied the modifications of the proposal, 18 master's students who used the initial version of the system and completed the *INT* problems participated in the *SUS* questionnaire. Their responses are summarized in Table 5.

Based on the *SUS* responses from 18 master's students before the modification, the average *SUS* score was 61.67, corresponding to *Grade D (OK)*. In terms of grade distribution, 6 students (33.33%) rated the system as *Grade C (Good)*, 10 students (55.56%) as *Grade D*

(OK), and 2 students (11.11%) as *Grade E (Poor)*, with no responses falling into *Grades A*, *B*, or *F*. This indicates that 66.67% of participants perceived the system's usability as below satisfactory (*Grades D–E*), while only 33.33% rated it positively (*Grade C*).

Table 5. Distribution of SUS responses (before).

Q	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Q1	0	1	4	10	3
Q2	4	7	4	1	2
Q3	1	1	3	8	5
Q4	3	4	2	6	3
Q5	1	2	0	10	5
Q6	5	6	5	1	1
Q7	0	0	6	7	5
Q8	2	1	6	7	2
Q9	0	0	4	8	6
Q10	1	1	3	9	4

Positive feedback was notably observed in Q1 (understanding *Flutter* concepts), Q3 (ease of use), Q5 (logical learning flow), and Q9 (documentation clarity), while Q2 (system complexity), Q6 (design inconsistency), and Q8 (documentation clarity) received more negative responses. These results suggest that, while the instructional content was generally helpful, key usability issues (especially in terms of navigation, design consistency, and clarity of feedback) limited the overall user experience.

6.4.4. Usability Result After Modification

After improving the interface and documentation of the proposal, we collected *SUS* responses from 16 fourth-year undergraduate students who used the modified system by solving the *INT* problems. Table 6 shows the distribution of their responses.

Table 6. Distribution of SUS responses (after).

Q	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Q1	1	2	4	6	3
Q2	5	5	4	2	0
Q3	0	2	5	4	5
Q4	4	2	5	4	1
Q5	0	2	3	7	4
Q6	7	3	3	2	1
Q7	0	0	2	10	4
Q8	9	4	2	1	0
Q9	0	1	1	8	6
Q10	6	6	3	1	0

Based on the *SUS* responses from 16 fourth-year undergraduate students after the system modification, the average *SUS* score was improved to 71.56, which corresponds to *Grade C (Good)* based on the *SUS* grade scale. As their details, the responses were distributed as follows: 4 students (25%) rated the system as *Grade A (Best Imaginable)*, 1 student (6.25%) as *Grade B (Excellent)*, 4 students (25%) as *Grade C (Good)*, and 7 students (43.75%) as *Grade D (OK)*. While a few participants reported very high usability ratings (*Grade A*), the overall average remained at 71.56 (*Grade C*), suggesting generally positive perceptions with notable areas for improvements.

This indicates that 56.25% of participants rated the system's usability as satisfactory (*Grades A–C*), while the remaining 43.75% rated it as *Grade D*, indicating that some aspects of the system could still be improved. Compared to the before-modification results, this reflects an overall improvement of +9.89 points, along with a shift toward higher satisfaction ratings. However, the presence of 43.75% of students rating *Grade D* indicates that further refinements of the system are needed.

Positive feedback was especially observed in Q1 (understanding *Flutter* concepts), Q3 (ease of use), Q5 (logical learning flow), Q7 (introductory documentation), and Q9 (clarity of core principles). In contrast, lower ratings were recorded in Q2 (system complexity) and Q6 (design inconsistency), indicating that some students experienced difficulties with interface navigations or perceived inconsistencies, which were recognized through student feedback and were addressed by implementing a feedback mechanism, as detailed in Section 7, particularly in the subsection *Feedback Mechanism Implementation*. Overall, the updated *SUS* scores reflect improved user satisfaction and usability, with the system moving into the "Good" category.

6.4.5. Statistical Comparison of SUS Scores

To verify the usability improvement of the proposed system, we compared the System Usability Scale (SUS) scores obtained in two evaluation rounds. The first round was conducted with 18 master's students using the initial version of the system, while the second round involved 16 fourth-year undergraduate students using the improved version system.

Table 7 shows the average SUS score and the standard deviation in each round. The results indicate that students in the second round gave higher and more consistent SUS scores compared to those in the first round, despite being more junior in academic level.

Table 7. SUS score comparison between two rounds.

Group	Participants	Mean SUS Score	Standard Deviation
Round 1 (Master)	18	61.67	10.85
Round 2 (Undergraduate)	16	71.56	14.63

An independent sample *t*-test was applied to the results of the two groups. It resulted in a *p*-value of 0.0350, which is below the standard 0.05 threshold. This suggests that the observed difference of the mean SUS scores is statistically significant. In addition, the Cohen's d effect size was calculated to be 0.7755, indicating a medium-to-large effect. These results confirm that the improved system delivered a noticeable enhancement in perceived usability.

Figure 6 compares the distribution of the SUS scores between the two rounds. The boxplot shows that students in Round 2 achieved higher median SUS scores with a wider

but still favorable spread. The statistical significance ($p = 0.0350$) and effect size (Cohen's d = 0.7755) are noted on the figure, confirming the effectiveness of the improved system in enhancing usability for beginner learners.

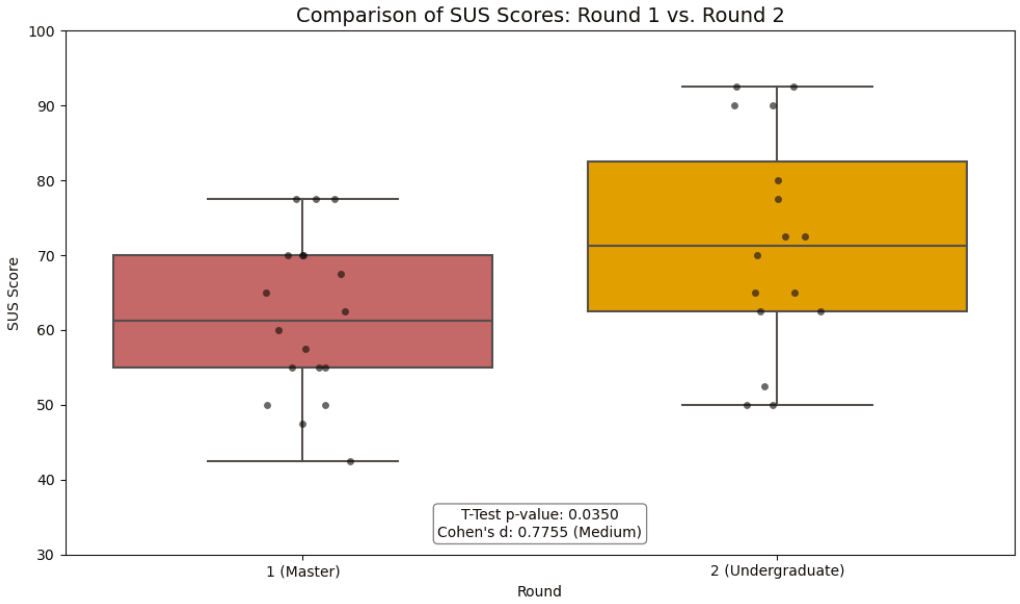


Figure 6. Comparison of SUS scores between two rounds with effect size (Cohen's d).

7. Free Opinions from Students

In this section, we analyze free opinions from students during two evaluation rounds to complement the *SUS* results.

7.1. First Round Evaluation

First, we discuss opinions from students during the first round.

7.1.1. Student Feedback

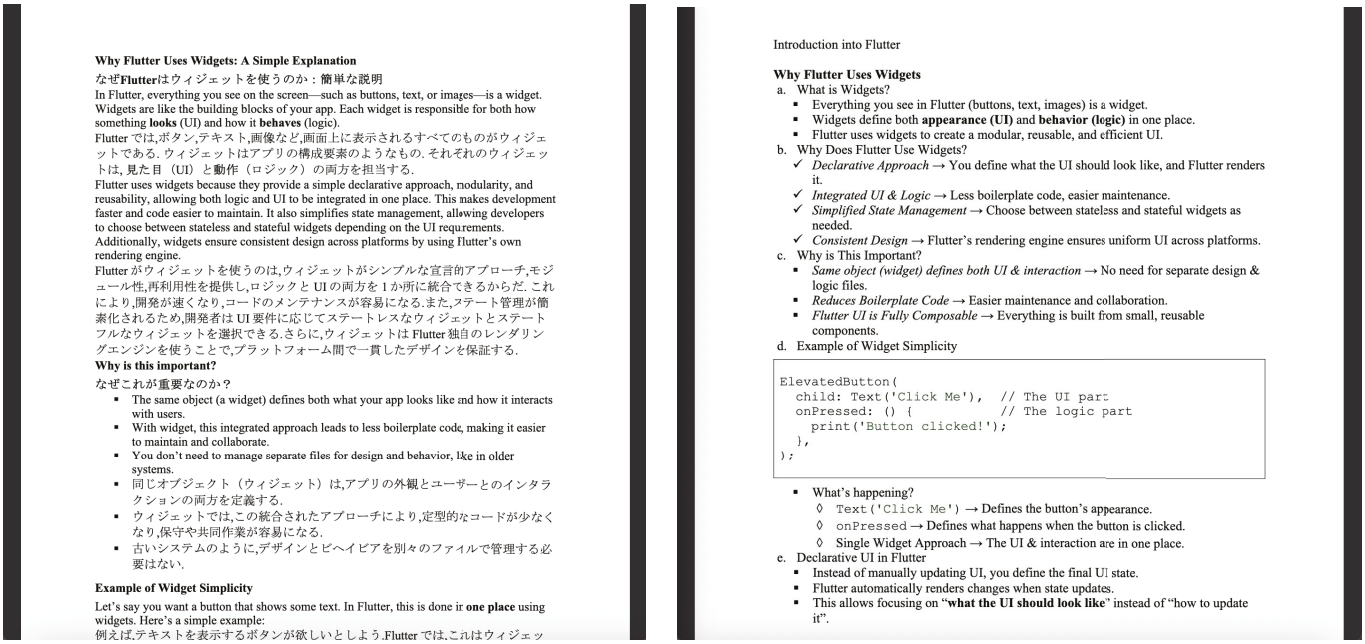
During the first round evaluation, the following opinions were collected:

- **System UI and Structure:**
  - There was no visual feedback when students submitted incorrect answers, which made it difficult to identify and fix mistakes;
  - Students found it inconvenient to look up relevant documentation manually and suggested displaying page references directly on the question screen;
  - The mixing of code snippets and problem descriptions within the same block reduced readability and caused confusion.
- **Documentation:**
  - The side-by-side bilingual format was considered distracting. Students preferred separate Japanese and English versions;
  - The difference in font size between headings and body text was too small, making it difficult to distinguish structure at a glance;
  - Explanations of similar topics (e.g., widget types) were scattered and should be grouped more clearly;
  - The relationships between core components such as `Class`, `Widget`, `Scaffold`, and `MaterialApp` were hard to grasp without a visual aid (diagram).

7.1.2. Implementation of the Modification

Based on the opinions in the first round, we implemented the following modifications to improve the documentation and user interface:

- The documentation was completely rewritten: Japanese and English versions were separated, the layout was improved, related topics were grouped, and long paragraphs were replaced with bullet points for better readability (as shown in Figure 7b compared to Figure 7a);
- Documentation references were added directly to each problem screen to guide students to relevant explanations without leaving the page (as shown in Figure 8);
- Several ambiguous question texts were rewritten to improve clarity and prevent misinterpretation (see revision in Figure 9b);
- The fill-in-blank answer fields were repositioned to align correctly with the context, improving the visual structure and usability (as shown in Figure 9b).



Why Flutter Uses Widgets: A Simple Explanation

なぜ**Flutter**はウィジェットを使うのか：簡単な説明

In Flutter, everything you see on the screen—such as buttons, text, or images—is a widget. Widgets are like the building blocks of your app. Each widget is responsible for both how something **looks** (UI) and how it **behaves** (logic).

Flutter では,ボタン,テキスト,画像など,画面上に表示されるすべてのものがウィジェットである. ウィジェットはアプリの構成要素のようなもの. それぞれのウィジェットは, 見た目（UI）と**動作**（ロジック）の両方を担当する.

Flutter uses widgets because they provide a simple declarative approach, modularity, and reusability, allowing both logic and UI to be integrated in one place. This makes development faster and code easier to maintain. It also simplifies state management, allowing developers to choose between stateless and stateful widgets depending on the UI requirements. Additionally, widgets ensure consistent design across platforms by using Flutter's own rendering engine.

Flutter がウィジェットを使うのは,ウィジェットがシンプルな宣言的アプローチ,モジュール性,再利用性を提供し,ロジックと UI の両方を 1 か所に統合できるからだ. これにより,開発が速くなり,コードのメンテナンスが容易になる.また,ステート管理が簡素化されるため,開発者は UI 要件に応じてステートレスなウィジェットとステートフルなウィジェットを選択できる.さらに,ウィジェットは Flutter 独自のレンダリングエンジンを使うことで,プラットフォーム間で一貫したデザインを保証する.

Why is this important?

なぜこれが重要なのか？

- The same object (a widget) defines both what your app looks like and how it interacts with users.
- With widget, this integrated approach leads to less boilerplate code, making it easier to maintain and collaborate.
- You don't need to manage separate files for design and behavior, like in older systems.
- 同じオブジェクト（ウィジェット）は,アプリの外観とユーザーとのインタラクションの両方を定義する.
- ウィジェットでは,この統合されたアプローチにより,定型的なコードが少なくなり,保守や共同作業が容易になる.
- 古いシステムのように,デザインとビヘイビアを別々のファイルで管理する必要はない.

Example of Widget Simplicity

Let's say you want a button that shows some text. In Flutter, this is done in **one place** using widgets. Here's a simple example:

例えば,テキストを表示するボタンが欲しいとしよう.Flutter では,これはウィジェッ

(**a**) Before.

Introduction into Flutter

Why Flutter Uses Widgets

a. What is Widgets?
   - Everything you see in Flutter (buttons, text, images) is a widget.
   - Widgets define both **appearance (UI)** and **behavior (logic)** in one place.
   - Flutter uses widgets to create a modular, reusable, and efficient UI.

b. Why Does Flutter Use Widgets?
   - ✓ *Declarative Approach* → You define what the UI should look like, and Flutter renders it.
   - ✓ *Integrated UI & Logic* → Less boilerplate code, easier maintenance.
   - ✓ *Simplified State Management* → Choose between stateless and stateful widgets as needed.
   - ✓ *Consistent Design* → Flutter's rendering engine ensures uniform UI across platforms.

c. Why is This Important?
   - *Same object (widget) defines both UI & interaction* → No need for separate design & logic files.
   - *Reduces Boilerplate Code* → Easier maintenance and collaboration.
   - *Flutter UI is Fully Composable* → Everything is built from small, reusable components.

d. Example of Widget Simplicity

```
ElevatedButton(
  child: Text('Click Me'),  // The UI part
  onPressed: () {           // The logic part
    print('Button clicked!');
  },
);
```

   - What's happening?
     - ◊ `Text('Click Me')` → Defines the button's appearance.
     - ◊ `onPressed` → Defines what happens when the button is clicked.
     - ◊ Single Widget Approach → The UI & interaction are in one place.

e. Declarative UI in Flutter
   - Instead of manually updating UI, you define the final UI state.
   - Flutter automatically renders changes when state updates.
   - This allows focusing on "**what the UI should look like**" instead of "how to update it".

(**b**) After.

Figure 7. Documentation improvements.

7.2. Second Round Evaluation

Next, we discuss opinions during the second round.

7.2.1. Students Feedback

In the second round, most of the students appreciated the improved structure and readability of the system. However, several participants still reported difficulties when their answers were incorrect, but we received no indications of what was wrong.

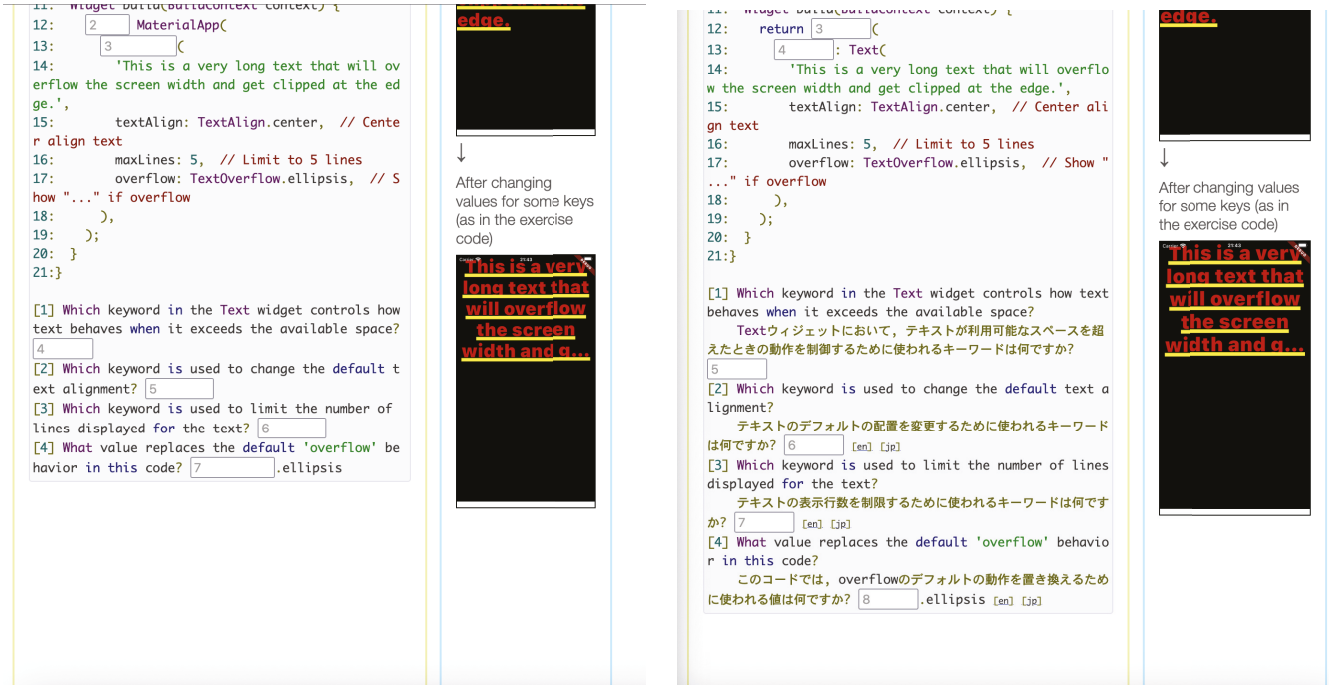
In particular, two students demonstrated significantly higher numbers of submission attempts. Upon further inquiry, they explained that the lack of feedback made it difficult to understand whether their mistakes were due to typographical errors, incorrect formatting, or misunderstandings of the programming concepts. This highlighted the need for a more informative feedback mechanism to help students refine their answers effectively.

Correspondence Between Each Topic and Related Flutter_INT Problems

※ Click the link to go reference pages directly

Flutter Concepts	Related Problem Numbers
Dart Operators	10, 11, 12, 16
Dart Type Data	10, 11, 12, 13, 14, 15, 16
Introduction to Widget	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
Key-value Format	3, 4, 5
Functions and Properties inside Widgets	6, 8, 12
Import package	1
main	1, 6, 7, 12, 14, 15
MyApp and Widget Tree	2, 4
Class and extends	2, 3, 6, 8, 9, 12, 15, 16
Override	2, 10, 16
Build Method	2, 3, 5, 10, 12, 14, 15, 16
BuildContext	2, 3, 5, 10, 12, 14, 15, 16
Return	2, 4, 5, 12, 13, 16
MaterialApp	3, 4, 5, 7, 8, 9, 10, 12, 16
Scaffold	5, 6, 7, 8, 9, 11, 12, 14, 15, 16
Stateless and Stateful	13, 15
setState	15
Versatile Function	14
Ternary Condition Operation	16
AppBar Widget	7, 11, 12, 13
Text Widget	4, 6, 7, 13
Container Widget	8
Row and Column Widget	9, 11, 14, 15, 16
Button Widget	10, 13

Figure 8. Instance IDs with direct link to corresponding documentation page.



(**a**) Before. (**b**) After.

Figure 9. Comparison of the question interface before and after the improvements.

7.2.2. Feedback Mechanism Implementation

To help students recognize the reasons behind their incorrect responses, we introduced a pop-up hint mechanism. It activates only when a non-empty answer is submitted, ensuring that feedback is given in response to actual input rather than empty fields.

We designed three types of hint logics:

1. **Case sensitivity alert:** If the normalized input matches the correct answer hash, the system assumes a case mismatch and displays a hint accordingly.

2. **Typo detection:** If normalization fails, the Levenshtein distance is calculated [30]. If the distance is 1 or less, a likely typo is reported.
3. **Predefined template hints:** For inputs that are significantly different, a task-specific hint is shown. These hints were generated beforehand using *ChatGPT-4o*, based on the structure and intent of each question, as explained next.

7.2.3. Template Hint Generation Using *ChatGPT*

In this study, we used *ChatGPT* to help generate hints that match the learning needs of beginners and fit the context of each question. Although we did not follow a formal prompt engineering method, our approach naturally included important techniques such as giving clear instructions, setting constraints, and refining prompts based on the results.

Specifically, we applied the following:

- **Instruction-based prompting**: Each prompt clearly explained the task (e.g., generating hints for fill-in-the-blank questions), as shown in Appendix A, Prompt 1;
- **Constraint specification**: We defined strict rules such as "do not include the correct answer", "avoid technical terms", and "reference the documentation", as shown in Appendix A, Prompt 2;
- **Iterative refinement**: We revised the prompts based on the initial outputs, requesting shorter, clearer, and more stylistically varied hints, as shown in Appendix A, Prompt 4;
- **Style control**: We selected appropriate hint types (e.g., Functionality Hint, Keyword Completion Hint, etc.) based on a style list suggested by the model, as shown in Appendix A, Prompt 1 and 2.

The actual generation process proceeded in four steps:

1. **Hint style selection**: We first asked *ChatGPT* to list possible hint styles suitable for coding exercises (Appendix A, Prompt 1). After reviewing the list, we selected seven styles that matched our educational goals, including functionality-based and documentation-referencing hints.
2. **Constraint-driven generation**: We instructed the model to generate hints based on our chosen styles, with clear constraints such as "no answer leakage" and "must be beginner-friendly" (Appendix A, Prompt 2).
3. **Feedback and revision**: Upon reviewing the output, we gave feedback on issues such as long sentences or uniformity in style, and requested improved versions (Appendix A, Prompt 4).
4. **Bilingual hint production**: The final set of hints was generated in both English and Japanese, and was integrated into our feedback pop-up system (Appendix A, Prompt 3).

Note that the integrated hint function shown here was implemented after the initial round of evaluation, and was not part of the system tested in Section 6. Its inclusion is based on user feedback and our system refinement efforts.

Examples of the actual hint messages generated by *ChatGPT* are shown in Figure 10, where the pop-up provides feedback based on the student's submitted answer. These hints follow the specified constraints (e.g., no direct answer, beginner-friendly language) and appear only when an incorrect answer is submitted.

Figure 10. Integrated feedback pop-up in the learning system.

8. Discussion

In this section, we analyze the evaluation results to assess the effectiveness, usability, and learning support of the proposed system, including the impact of the feedback mechanism.

8.1. Interpretation and Answers to Research Questions

We examine whether the integration of a documentation, a source code, a visual output, and *INT* problems have improved conceptual understanding and the system usability, from the results in two evaluation rounds.

In RQ1, the high average correct answer rates observed in both evaluations suggest that the students were able to understand and apply *Flutter/Dart* programming concepts effectively using the proposed system. In the first round, 17 out of 23 students achieved perfect scores, and in the second round, 21 out of 25 students did so. The consistent correctness, even after modifications, indicates that the integration of documentation and interactive elements helped reinforce students' conceptual learning.

For RQ2, the average SUS score increased from 61.67 in the first round to 71.56 in the second round. To verify whether this difference was statistically significant, we conducted an independent sample t-test, resulting in a p-value of 0.035, which is below the 0.05 threshold. This indicates a statistically significant improvement in system usability. The effect size (Cohen's $d = 0.79$) suggests a medium-to-large effect. These findings imply that modifications such as improved documentation structure, clearer instructions, and interface enhancements contributed meaningfully to a more satisfactory and intuitive learning experience.

These findings demonstrate that the proposed learning system effectively supports both the learning and usability goals.

8.2. Limitations

Our study in this paper has several limitations that should be acknowledged.

First, the problem instance creation relied on manual authoring decisions, although the topics and blank selections were carefully designed following the documentation contents and grammar-focused principles for the *GUP* format. Future works should focus on standardizing the blank selection process to improve the accuracy, fairness, and scalability. For example, we will define rule-based criteria tied to syntax type, widget category, or difficulty levels so that instance generations become further consistent and automated across problems.

Second, the evaluation was conducted with students sharing a similar background, where the number of participants and institutional scope were limited. They have prior experiences in *object-oriented programming (OOP)* (e.g., *Java*, *JavaScript*, *C++*, or *Python*) but have no prior knowledge of *Flutter/Dart*. This alignment supports the system's intended learner profile. However, the use of two distinct participant groups, where one group consists of master's students and another of fourth-year undergraduate students, may introduce variability due to differences in academic levels, motivations, or prior exposures to related concepts between them. Future evaluations will be necessary with more controlled or homogeneous participant groups to improve the validity and comparability of the results.

Third, the evaluation was conducted in a short-term setting, involving detailed analysis and active feedback collection from participants. The evaluation of long-term retention or knowledge transfer beyond the immediate learning sessions was not conducted. Evaluating these aspects remains an important direction for future works.

Fourth, this study did not incorporate pre/post testing or baseline comparison groups such as documentation-only or conventional material conditions. As an initial exploratory implementation to students in real educational settings with formal classes, our evaluation prioritized practical usage and feedback collections rather than formal measurement of learning gains.

While the high correct answer rates indicate successful completions of the tasks by students, they may not fully capture their deeper conceptual understanding or transferability. Without follow-up assessments, it remains possible that some students relied on answer pattern recognitions rather than genuine comprehensions of the concepts. Evaluating these aspects remains an important direction for future works.

8.3. Implications and Future Works

The results of this study offer several implications on the design of programming learning environments, particularly for introductory *Flutter/Dart* programming. The proposed integration of beginner-oriented documentation with two exercise problem formats in *INT* demonstrated that the structured, context-aware guidance can significantly enhance understanding of introductory *Flutter/Dart* programming, especially in independent study environments.

The combination of a source code, a visual output through a screenshot, and an interactive solving process implemented in the *INT* format have helped students connect the abstract syntax with the concrete interface behavior. This approach is particularly valuable when learning UI-focused frameworks such as *Flutter*. Furthermore, the grammar-focused learning design of *GUP* and the structured completion tasks of *EFP* proved effective in bridging the gap between students' prior *object-oriented programming (OOP)* experiences and the *Flutter*'s declarative model.

Based on these findings, several directions are planned for future developments. First, we aim to expand the set of *INT* problems by covering more advanced *Flutter/Dart* topics, such as state management patterns and widget compositions. Alongside this, we will explore automating the blank selection process for creating a new *INT* instance

using a rule-based method, reducing manual workloads and ensuring consistency in problem generations.

Second, we plan to enhance the system's adaptability by incorporating features that respond to individual learning progress or error patterns. In addition, we will explore the development of a mobile application of the system using *Flutter/Dart* to promote greater accessibility and enable students to engage in *Flutter/Dart* learning at any time and any place.

Furthermore, we intend to adopt more rigorous evaluation designs in future studies, including the use of pre/post testing and baseline comparison groups, to more accurately assess conceptual learning gains and further validate the educational effectiveness of the proposed system.

In particular, we aim to include methods for assessing conceptual transfers, such as follow-up tasks of writing new codes or explaining learned concepts. To improve internal validity of evaluations, more homogeneous participant groups (e.g., students of the same academic level and background) will be considered to reduce potential confounding factors when comparing different evaluation rounds.

9. Conclusions

This paper presented the *Integrated Introductory Problem (INT)* format for the *Flutter Programming Learning Assistant System (FPLAS)*, designed to support beginner-oriented self-study in *Flutter/Dart cross-platform mobile application* programming. By integrating a beginner-friendly documentation, a source code, a visual output, and exercises, the proposal helps students bridge the gap between abstract syntax and interface behavior. For evaluations, 16 *INT* instances were generated and tested in two rounds with 48 students in total at Okayama University, Japan. In both rounds, high answer correctness rates were achieved, and improvements in *System Usability Scale (SUS)* scores were observed in the second round after refining the documentation and interface. The results indicate that the proposed learning system is effective in helping novice students understand the core principles of *Flutter/Dart* programming, even without prior experience. In future works, we aim to extend the coverage of *INT* to include more advanced topics such as state management and widget composition. We also plan to enhance the feedback function and implement adaptive guidance features, and automate the problem generation procedure using a rule-based method. Additionally, a mobile application version of the platform will be developed to support flexible, anytime learning for cross-platform mobile development education.

Supplementary Materials: The following supporting information can be downloaded at: https://github.com/adinekinar/int-plas-2025 (accessed on 20 September 2025).

Author Contributions: Conceptualization, S.A.K. and N.F.; Methodology, S.A.K.; Software, S.A.K., S.T.A., and H.H.S.K.; Investigation, S.A.K.; Writing—original draft preparation, S.A.K.; Writing—review and editing, S.A.K. and N.F.; Supervision, N.F.; Project administration, N.F.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The ethical review and approval were waived for this study by the Ethics Committee of Okayama University because it utilized only fully anonymized data collected from regular classroom activities, posing no risk to participants.

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

Data Availability Statement: The original contributions presented in the study are included in the article (and Supplementary Materials).

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

Appendix A. Prompt Examples Used for Hint Generation

Below are excerpts of the actual prompts used during the template hint generation process:

- **Prompt 1 – Style Discovery:**

> I need you to help me provide hints for students to solve my *Flutter* task (especially for fill-in-the-blank code and *Flutter/Dart* keyword understanding). Firstly, I want to know what kind of hint style type that we can use. Could you make a list for me?

- **Prompt 2 – Style Selection and Constraints:**

> Thank you for providing the list. After reading it, it seems that the style of type hint that is suitable for the task I have given is:
>
> 1. Functionality Hint
> 2. Syntax/Form Hint
> 3. Keyword Completion Hint
> 4. Code Role Identification
> 5. Multiple-Choice Elimination Guide
> 6. Hint Question
> 7. Documentation Reference Hint (this is highly recommended)
>
> Since I have already selected the hint types, can you try creating hints if I provide the questions? If necessary, I can also provide the answers.
> However, you must ensure that the hints always follow the rules below:
>
> – Do not give away the answer directly;
> – Do not include the answer in the hint sentence at all;
> – Do not use overly technical terminology;
> – Make it easy for laypeople to understand;
> – Indicate the page and section (or subsection) in the documentation that we created earlier.

- **Prompt 3 – Bilingual Hint Request:**

> I will provide you with the question set and documentation. Please create hints in English and Japanese versions.

- **Prompt 4 – Revision Feedback:**

> Here is my feedback on the hints you created:
>
> – The sentences in the hints you created are too long;
> – You can use several different styles of hints for different questions.
>
> Please correct them.

References

1. Statista. Smartphone Users Worldwide—Forecast to 2029. Available online: https://www.statista.com/forecasts/1143723/smartphone-users-in-the-world (accessed on 11 August 2025).

2. Pinto, C.M.; Coutinho, C. From native to cross-platform hybrid development. In Proceedings of the 2018 International Conference on Intelligent Systems (IS), Funchal, Portugal, 25–27 September 2018; pp. 669–676.
3. Jošt, G.; Taneski, V. State-of-the-art cross-platform mobile application development frameworks: A comparative study of market and developer trends. *Informatics* **2025**, *12*, 45. [CrossRef]
4. Martinez, M.; Lecomte, S. Towards the Quality Improvement of Cross-Platform Mobile Applications. In Proceedings of the 2017 IEEE/ACM 4th International Conference on Mobile Software Engineering and Systems (MOBILESoft), Buenos Aires, Argentina, 20–21 May 2017; pp. 184–188.
5. Zou, D.; Darus, M.Y. A Comparative Analysis of Cross-Platform Mobile Development Frameworks. In Proceedings of the 2024 IEEE 6th Symposium on Computers & Informatics (ISCI), Kuala Lumpur, Malaysia, 10 August 2024; pp. 84–90.
6. You, D.; Hu, M. A comparative study of cross-platform mobile application development. In Proceedings of the 12th Annual Conference of Computing and Information Technology Research and Education New Zealand (CITRENZ 2021), Wellington, New Zealand, 14–16 July 2021.
7. Gülcüoğlu, E.; Üstün, A.B.; Seyhan, N. Comparison of Flutter and React native platforms. *J. Int. Appl. Manag.* **2021**, *12*, 129–143. [CrossRef]
8. Dart. Dart Overview. Available online: https://dart.dev/overview (accessed on 27 June 2025).
9. Flutter. Build Apps for Any Screen. Available online: https://flutter.dev/ (accessed on 27 June 2025).
10. Bailey, T.; Biessek, A. *Flutter for Beginners: An Introductory Guide to Building Cross-Platform Mobile Applications with Flutter 2 and Dart*; Packt Publishing: Birmingham, UK, 2021.
11. Patta, A.R.; Funabiki, N.; Lu, X.; Syaifudin, Y.W. A study of grammar-concept understanding problem for Flutter cross-platform mobile programming learning. In Proceedings of the International Conference on Vocational Educational and Electrical Engineering (ICVEE), Surabaya, Indonesia, 14–15 October 2023; pp. 249–254.
12. Aung, S.T.; Funabiki, N.; Aung, L.H.; Kinari, S.A.; Mentari, M.; Wai, K.H. A study of learning environment for initiating Flutter app development using Docker. *Information* **2024**, *15*, 191. [CrossRef]
13. Kinari, S.A.; Funabiki, N.; Aung, S.T.; Wai, K.H.; Mentari, M.; Puspitaningayu, P. An independent learning system for Flutter cross-platform mobile programming with code modification problems. *Information* **2024**, *15*, 614. [CrossRef]
14. Kinari, S.A.; Funabiki, N.; Aung, S.T.; Mentari, M.; Puspitaningayu, P. A study of guide documentation for introductory Flutter programming learning with exercises. In Proceedings of the International Conference on Consumer Technology (ICCT-Pacific), Shimane, Japan, 29–31 March 2025; pp. 1–4.
15. Brooke, J. SUS: A quick and dirty usability scale. *Usability Evaluation in Industry*; Jordan, P.W., Thomas, B., Weerdmeester, B., McClelland, I.L., Eds.; Taylor & Francis: London, UK, 1995; p. 189.
16. Nagaraj, N.K.; Prabakaran, B.; Ramkumar, M.O. Application development for a project using Flutter. In Proceedings of the International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 20–22 October 2022; pp. 947–950.
17. Zou, D.; Bin Darus, M.Y.; Ramli, A.B. Investigating developer experiences with UI components in Flutter: Challenges and implications. In Proceedings of the International Visualization, Informatics and Technology Conference (IVIT), Kuala Lumpur, Malaysia, 7–8 August 2024; pp. 128–131.
18. Boukhary, S.; Colmenares, E. A clean approach to Flutter development through the Flutter clean architecture package. In Proceedings of the International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 5–7 December 2019; pp. 1115–1118.
19. Radadiya, S.; Ramchandani, S. The impact of Flutter development. *Int. Res. J. Mod. Eng. Technol. Sci.* **2025**, *7*. [CrossRef]
20. Flutter. Flutter Architectural Overview. Official Documentation, May 2025. Available online: https://docs.flutter.dev/resources/architectural-overview (accessed on 12 August 2025).
21. Sharma, S.; Khare, S.; Uniyal, V.; Verma, S. Hybrid development in Flutter and its Widgets. In Proceedings of the International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 6–7 October 2022.
22. Wang, Y.-L.; Wang, C.-Y.; Hsiao, H.-S. The CodingHere platform for programming courses: Improving learning motivation, effectiveness, and cognitive load. *Electronics* **2022**, *11*, 1680.
23. Hinrichs, T.; Burau, H.; von Pilgrim, J.; Schmolitzky, A. A scalable online programming platform for software engineering education. In *Software Engineering 2021 Satellite Events*; Götz, S., Linsbauer, L., Schaefer, I., Wortmann, A., Eds.; Lecture Notes in Informatics (LNI); Gesellschaft für Informatik: Bonn, Germany, 2021.
24. Combéfis, S.; Girgis, M.; Hage, J.; Ihantola, P.; Leinonen, J.; Rubio, M.A. Automated code assessment tools: A review of the state of the art. *Software* **2022**, *1*, 144–159. [CrossRef]
25. Jevtić, M.; Slijepčević, S.; Tomić, A.; Vučetić, M. Source code analysis in programming education: Evaluating learning content with self-organizing maps. *Appl. Sci.* **2023**, *13*, 5719. [CrossRef]
26. Hsueh, C.-H.; Wu, J.-C.; Yang, S.-T.; Liu, C.-C.; Lin, K.-H. Design of an online programming platform and a study on learners' testing ability. *Electronics* **2023**, *12*, 4596. [CrossRef]

27. Hasan, M.A.; Ahsan, M. Automated assessment in mobile programming courses: An evaluation system for Java and Kotlin apps. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 295–302.
28. Togni, J. Development of an inclusive educational platform using open technologies and machine learning: A case study on accessibility enhancement. *arXiv* **2025**, arXiv:2503.15501v1.
29. Renkl, A. The Worked-Out Examples Principle in Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning*, 2nd ed.; Mayer, R.E., Ed.; Cambridge University Press: New York, NY, USA, 2014; pp. 391–412.
30. Levenshtein, V.I. Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics* **1966**, *10*, 707–710.

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.

computers

MDPI

Article

The Learning Style Decoder: FSLSM-Guided Behavior Mapping Meets Deep Neural Prediction in LMS Settings

Athanasios Angeioplastis [1], John Aliprantis [2], Markos Konstantakis [2,*], Dimitrios Varsamis [1] and Alkiviadis Tsimpiris [1]

[1] Department of Computer, Informatics and Telecommunications Engineering, International Hellenic University, 62124 Serres, Greece; aagiop@gmail.com (A.A.); dvarsam@ihu.gr (D.V.); atsimpiris@ihu.gr (A.T.)
[2] Department of Cultural Technology and Communication, Aegean University, 81100 Mytilene, Greece; jalip@aegean.gr
* Correspondence: mkonstadakis@aegean.gr

Abstract: Personalized learning environments increasingly rely on learner modeling techniques that integrate both explicit and implicit data sources. This study introduces a hybrid profiling methodology that combines psychometric data from an extended Felder–Silverman Learning Style Model (FSLSM) questionnaire with behavioral analytics derived from Moodle Learning Management System interaction logs. A structured mapping process was employed to associate over 200 unique log event types with FSLSM cognitive dimensions, enabling dynamic, behavior-driven learner profiles. Experiments were conducted across three datasets: a university dataset from the International Hellenic University, a public dataset from Kaggle, and a combined dataset totaling over 7 million log entries. Deep learning models including a Sequential Neural Network, BiLSTM, and a pretrained MLSTM-FCN were trained to predict student performance across regression and classification tasks. Results indicate moderate predictive validity: binary classification achieved practical, albeit imperfect accuracy, while three-class and regression tasks performed close to baseline levels. These findings highlight both the potential and the current constraints of log-based learner modeling. The contribution of this work lies in providing a reproducible integration framework and pipeline that can be applied across datasets, offering a realistic foundation for further exploration of scalable, data-driven personalization.

Keywords: Learning Style Modeling; personalized learning; LMS behavior analytics; Felder–Silverman Model (FSLSM); adaptive learning systems; educational data mining; Moodle personalization; behavior-based profiling

1. Introduction

In the era of digital transformation in education, the need for personalized learning has emerged as a critical factor to enhance the effectiveness of the learning process. Understanding individual differences between students, particularly learning styles, constitutes a fundamental aspect of personalized education. The Felder–Silverman Learning Style Model (FSLSM) is a widely recognized framework that categorizes learning preferences in four dimensions, offering a comprehensive approach to understanding how learners process, perceive, receive, and comprehend information [1–3].

Learning Management Systems (LMSs), such as Moodle, have become significant sources of data regarding student behavior in online learning environments [4,5]. The Moodle-generated logs for each user action store valuable information about learners'

activities and preferences. Analyzing these data can provide invaluable information on student learning styles and support the provision of personalized and customized learning experiences [6–8].

Traditionally, the identification of learning styles has been based on self-assessment questionnaires. However, contemporary approaches seek to exploit automated methods based on technologies such as data mining and artificial intelligence to enable the implicit detection of learning styles through the analysis of LMS log files. These methods aim to uncover patterns, preferences, and behavioral trends among students within the learning platform [9,10].

In this study, we propose a hybrid learner profiling methodology that integrates explicit psychometric profiling (via an extended FSLSM questionnaire) with behavioral analytics extracted from LMS interaction logs. A structured mapping framework translates raw log events into cognitive learning style dimensions, thereby supporting the construction of dynamic, behavior-driven learner profiles. To evaluate the approach, we conducted experiments using three datasets: (i) a real-world dataset from the International Hellenic University (IHU), containing over 4.3 million Moodle log entries; (ii) a public dataset from Kaggle, representing a different institutional context; and (iii) a merged dataset combining the two, totaling over 7 million interaction records. We trained three deep learning architectures (a feedforward neural network, a BiLSTM, and a pretrained MLSTM-FCN) on regression, binary classification, and multi-class grade prediction.

The results highlight both opportunities and limitations. While binary classification achieved moderate accuracy that could inform practical applications such as early alerts, regression and multi-class tasks performed close to baseline, underscoring the difficulty of predicting fine-grained outcomes solely from behavioral logs. Rather than demonstrating superior performance, the contribution of this work lies in offering a transparent integration framework, an extensible pipeline applicable to new datasets, and a critical examination of the potential and limits of FSLSM-informed behavioral profiling.

The remainder of this paper is structured as follows: Section 2 introduces the theoretical foundations of the Felder–Silverman Learning Style Model (FSLSM), which underpins the learner profiling framework. Section 3 reviews prior work in learner modeling and behavioral analytics in educational settings. Section 4 presents the proposed methodology for mapping LMS interaction data to cognitive learning style dimensions. Section 5 describes the datasets, prediction tasks, and the deep learning architectures employed. Section 6 reports the experimental findings across regression and classification tasks, while Section 7 presents the learning outcomes. Furthermore, Section 8 interprets the results in light of pedagogical theory and system design considerations. Finally, Section 9 summarizes the key contributions and outlines directions for future research.

2. Felder–Silverman Learning Style Model (FSLSM)

The Felder–Silverman Learning Style Model (FSLSM) is a widely adopted framework for describing individual learning preferences across four cognitive dimensions: processing (active vs. reflective), perception (sensing vs. intuitive), reception (visual vs. verbal), and understanding (sequential vs. reflective) [11–13]. It was originally developed for engineering education but has since been applied extensively in personalized learning systems [2,14,15].

Learner profiling under FSLSM is typically performed using the Index of Learning Styles (ILS) questionnaire, a 44-item self-assessment tool that classifies learners' preferences by degree (balanced, moderate, or strong). The resulting profiles support pedagogical adaptation, including content delivery and interface design [16,17].

Beyond self-reported styles, recent research has emphasized behavioral profiling through LMS analytics. Studies have mapped user actions such as quiz activity, forum engagement, and resource navigation to FSLSM dimensions, enabling dynamic, data-driven personalization [7,18].

FSLSM-based personalization has been integrated into various recommendation systems using content-based, collaborative, and hybrid filtering techniques, as well as machine learning approaches such as decision trees, SVMs, and neural networks. These systems aim to align learning resources with cognitive preferences, often achieving improvements in engagement and performance [19,20].

Overall, FSLSM remains a foundational model in adaptive learning research. When combined with behavioral analytics, it enables intelligent systems to deliver more tailored and effective learning experiences in LMS-based environments.

3. Related Work

Recent years have witnessed a growing interest in developing personalized e-learning environments by leveraging learning style models, behavioral data, and recommendation technologies. A variety of studies have proposed frameworks that detect, model, or utilize students' learning styles, often based on the Felder–Silverman Learning Style Model (FSLSM), to enhance learning outcomes and adapt educational content [14,15,17,21].

Nazempour and Darabi [9] proposed a behavior-based personalized learning approach in virtual environments, using data mining techniques to extract meaningful learning patterns from LMS log data. Their work emphasizes the potential of behavioral analytics in implicitly detecting learning preferences and improving adaptive learning strategies. Murtaza et al. [10] addressed the challenges of AI-based personalized e-learning systems by proposing a framework that integrates multiple machine learning techniques for real-time adaptation. Their approach focuses on scalability and robustness, particularly emphasizing the dynamic modeling of user learning profiles.

Zlatkovic et al. [22] adjusted the FSLSM within Moodle-based adaptive e-learning systems. They developed a method to calibrate learning object delivery to match the learning styles inferred from FSLSM dimensions, aiming to improve learner engagement, while Zagulova et al. [23] conducted an empirical analysis linking FSLSM-based learner profiles to academic performance. Their findings suggest that personalization based on learning styles can have a measurable impact on student success, particularly in e-learning contexts.

Furthermore, Fernando-Raguro et al. [18] explored the extraction of student engagement patterns from LMS logs using decision trees and K-means clustering. Their work demonstrates the feasibility of deriving learner behavior patterns and grouping students accordingly to personalize learning interventions. Kaur et al. [24] developed an intelligent profiling system within an e-learning environment aimed at enhancing user adaptation. Their system used behavioral indicators combined with clustering methods to build adaptive learning pathways.

Valencia et al. [11] proposed the classification of learning styles based on Bayesian Networks, directly utilizing responses from the Index of Learning Styles (ILS) questionnaire. Their approach highlighted the potential of probabilistic models in accurately predicting learner profiles. Also, Hasibuan et al. [19] designed a recommendation model for learning materials tailored to students' FSLSM profiles. Their system maps learning resources to specific FSLSM dimensions and recommends content accordingly, achieving notable improvements in user satisfaction.

Ezzaim et al. [6,7] conducted a systematic review on AI-based learning style detection methods. Furthermore, in their implementation study [14], they developed a machine

learning-based multi-factor adaptive system that combines learning styles with performance metrics to drive content recommendation.

Moreover, Ait Daoud et al. [2] applied FSLSM-based analysis to student interaction data within a blended-learning Moodle course. They examined how different FSLSM dimensions correlate with students' engagement and performance in a real educational setting. Nafea et al. [17] proposed a learning object recommendation framework grounded in FSLSM dimensions. They integrated collaborative filtering and content-based filtering strategies, aiming to improve recommendation accuracy through hybridization.

Prabpal and Nitiwatthana [20] enhanced the detection of learning styles using advanced machine learning techniques, proposing improved feature engineering approaches for higher classification accuracy based on behavioral data. Buciuman and Potra [16] explored the integration of eye-tracking data with AI models to support personalized learning environments. Their study proposed that multimodal data sources could further refine learning style detection and adaptive interventions.

Hwang et al. [1] investigated the necessity of adaptive learning systems by empirically examining learning outcomes across different learning style groups. Their work validated the pedagogical significance of adapting instruction to learning style preferences.

Finally, in our previous work [8], we developed a predictive model for student performance using educational data mining techniques. While their primary focus was on performance prediction, their study reinforces the importance of fine-grained learner modeling [25–27].

4. Personalization Process

As highlighted in the related work (Section 3), significant progress has been made in utilizing learning style models and behavioral analytics to personalize educational experiences. However, most existing approaches either focus narrowly on recommending learning materials or rely predominantly on static learner profiling. There remains a need for methodologies that can dynamically map learner behaviors to cognitive styles, enabling the creation of truly personalized knowledge pathways tailored to each learner's capabilities.

Motivated by this gap, our study aims to design and evaluate a hybrid framework that combines explicit learning style profiling with behavioral indicators extracted from LMS interactions. By doing so, we seek to build detailed learner models that go beyond surface-level personalization and facilitate individualized knowledge construction based on students' preferred cognitive processing modes. This approach aspires not only to recommend appropriate content but also to support learners in navigating educational environments more effectively according to their intrinsic strengths and needs.

In the following sections, we describe the overall profiling framework, the mapping of LMS events to FSLSM dimensions, and the data analysis procedures adopted in our study.

4.1. Learner Profiles Based on FSLSM

Following FSLSM [2,16], learners are categorized across four main dimensions: active/reflective, sensing/intuitive, visual/verbal, and sequential/global. Each dimension reflects a specific cognitive preference that influences how learners interact with educational materials. Based on prior work [2,9,28,29], we define eight detailed learner profiles: Active, Reflective, Sensing, Intuitive, Visual, Verbal, Sequential, and Global.

Each profile is characterized by preferred interaction types:

- Active learners favor direct engagement (e.g., quizzes, assignments).
- Reflective learners benefit from exploratory, discussion-based activities.
- Sensing learners engage with concrete and factual content (e.g., multimedia).
- Intuitive learners prefer abstract and conceptual material.

- Visual learners excel with diagrams, videos, and visual aids.
- Verbal learners thrive with text-based or linguistic tasks (e.g., forums).
- Sequential learners appreciate linear, step-by-step learning.
- Global learners prefer holistic, big-picture approaches.

A graphical overview of the FSLSM learner profiles and their preferred interaction types is presented in Figure 1. This summary diagram maps each of the four FSLSM dimensions to their corresponding preferred modalities, making it clear how specific learner types engage best with different content formats. By visually clustering profiles and interaction modes, the figure highlights opportunities for targeted instructional design.

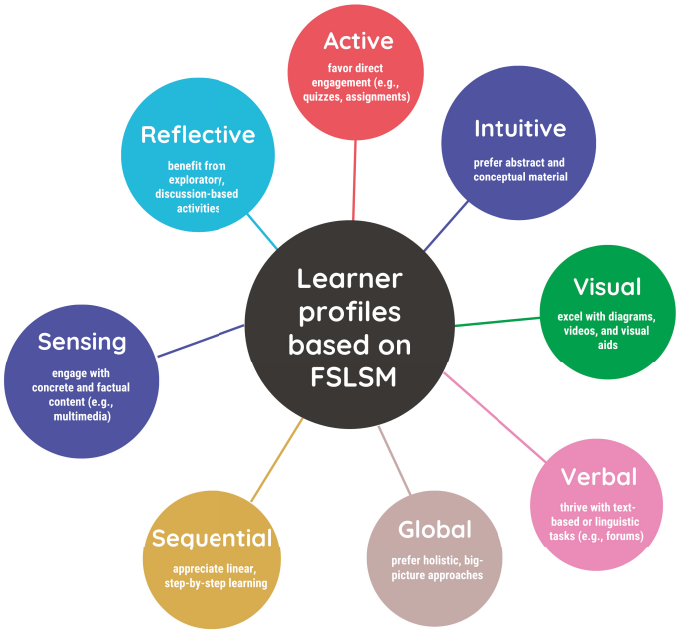


Figure 1. Overview of learner profiles based on FSLSM and their preferred interaction types.

4.2. Mapping LMS Interaction Logs to Learner Profiles

To effectively construct dynamic learner profiles, it is critical to accurately interpret user interaction behaviors recorded within the LMS environment. In this study, Moodle platform logs were analyzed, capturing diverse types of student actions, such as quiz attempts, forum posts, content views, assignment submissions, and time-on-task metrics.

The primary dataset utilized was extracted from Moodle's mdl_logstore_standard_log table. Specifically, we used the data set from the Department of Computer, Informatics and Telecommunications Engineering at the International Hellenic University (IHU) in Serres, Greece, featuring over 60 courses across 10 semesters for its undergraduate program, and log info of around 6000 students. This table systematically records every significant student interaction occurring within the platform. Each log entry includes, among other things, four critical attributes that we focus on to create a student's behavioral patterns:

- log_eventname: A textual description of the event type (e.g., "Course viewed", "Quiz attempted");
- log_action: The action associated with the event (e.g., "create", "view", "update");
- log_target: The object or target that the event concerns (e.g., "course", "quiz", "assignment");
- log_objecttable: The database table associated with the object (e.g., course, quiz_attempts).

To align behavioral data with FSLSM learning profiles, we first generated all possible combinations of four key behavioral indicators from the Moodle logs, resulting in over

200 unique interaction types. Each combination represented a distinct learning behavior, which we contextualized using the official Moodle documentation.

For each unique pattern, we assigned a descriptive label (e.g., "Assignment submission," "Quiz viewed") and grouped it into broader pedagogical categories such as content engagement, assessment, or collaboration. This step translated raw system logs into meaningful educational actions.

Next, we mapped each event type to FSLSM dimensions using a structured three-phase approach. First, actions were classified by pedagogical intent, e.g., active (e.g., quiz submission) versus passive (e.g., resource viewing). Then, these were aligned with relevant FSLSM dimensions based on prior literature [2,9] and Moodle-specific behaviors. Finally, a scoring system (+1, 0, –1) was applied to indicate alignment strength, and scores were normalized across learners to enable comparisons.

This mapping process enabled the creation of dynamic, behavior-based FSLSM profiles that complement traditional questionnaire-based assessments, allowing for more robust and adaptive learner modeling. The complete event-to-FSLSM mapping used in this study is presented in Appendix A. This array includes more than 200 interaction types and their corresponding cognitive dimension scores and serves as the core component of our profiling engine.

4.3. Data Analysis Process

The scoring schema was implemented as a lookup procedure using the mapping table provided in Appendix A, ensuring full reproducibility. The behavioral mapping process follows a structured six-step pipeline, enabling the transformation of LMS interaction logs into cognitively meaningful FSLSM learner profiles. The steps are outlined below.

- **Event Classification:**
  - LMS log entries are reviewed and labeled based on their pedagogical function.
  - Events are grouped into categories such as active participation (e.g., quiz submissions), reflective exploration (e.g., forum reading), or sensory engagement (e.g., video watching).
- **FSLSM Dimension Mapping:**
  - Each event category is mapped to one or more FSLSM dimensions (e.g., Active, Visual, Sequential).
  - This mapping is based on empirical literature and domain-specific pedagogical interpretations.
- **Weight Assignment:**
  - A numeric score is assigned to each event-dimension pair:
    - $+1$ for strong positive alignment;
    - 0 for neutral/no clear association;
    - -1 for negative/misaligned behavior.
- **Profile Construction:**
  - Learners' behavioral scores are aggregated across all logged interactions.
  - The result is a multi-dimensional cognitive profile vector reflecting FSLSM-aligned behavior.
- **Cross-Validation with Questionnaire Data:**
  - Behavioral profiles are compared to the results of the FSLSM Index of Learning Styles (ILS) questionnaire.
  - Any mismatches between observed behavior and self-reported preferences are flagged for interpretation.

- **Normalization and Interpretation:**
  - Final scores are normalized to a common scale across all learners.
  - Profiles are interpreted to reveal dominant or moderate tendencies per FSLSM dimension.
  - This step supports adaptive personalization strategies and learner clustering.

This structured procedure bridges the gap between static questionnaire assessments and dynamic behavioral observations, enabling more robust personalization in LMS environments.

4.4. Example Analysis of Learner Profiles

To demonstrate the implementation of our methodology, we present an indicative learner profile derived from actual LMS interaction logs, as shown in Table 1. The dataset was extracted from a Moodle-based environment and pre-processed to identify behavioral patterns across multiple interaction types, including resource views, quiz submissions, forum activity, and content navigation.

Table 1. Example learner profile mapping based on LMS interaction logs.

Profile	Score	Interpretation
Active	2200	Very high engagement with action-oriented activities such as quizzes, assignments, and forum posting.
Reflective	−288	Low tolerance for open-ended or unstructured learning pathways; struggles with exploratory navigation.
Sensor	1700	Strong preference for concrete, sensory-rich content (e.g., factual resources, examples).
Intuitive	174	Limited engagement with abstract material or conceptual frameworks; prefers practical over theoretical tasks.
Visual	1470	Frequent access to visual learning resources such as diagrams, video lectures, or presentations.
Verbal	999	Moderate interaction with textual content (e.g., PDFs, readings, discussions).
Sequential	2268	Consistent adherence to structured, linear navigation patterns; prefers step-by-step learning.
Global	532	Some ability to integrate content holistically, but not dominant; likely favors localized rather than big-picture reasoning.

Each log event was classified and assigned a score based on its pedagogical intent, then mapped to the appropriate FSLSM dimension using the weighted scheme outlined in Section 4.3. Cumulative scores were computed per dimension, providing a multi-dimensional behavioral representation of the student.

The input data for this analysis consisted of raw Moodle logs containing event-level metadata (e.g., event type, resource ID, timestamp, user ID). Each event was tagged and weighted according to its correspondence to FSLSM dimensions, as explained in Section 4.3. After aggregation and normalization, the scores reflected the degree to which the student's interactions aligned with each learning style.

This fine-grained behavioral profiling enables targeted personalization strategies. For example, a student scoring high on *Active* and *Visual* dimensions could benefit from interactive simulations and video content, while another with stronger *Verbal* and *Intuitive* tendencies might receive in-depth reading materials and conceptual explanations. Furthermore, scores such as *Reflective* (a derived metric capturing unstructured navigation

behavior) provide insight into students' tolerance for non-linear learning experiences, supporting dynamic adjustment of content flow.

The complete dataset with all unique interaction mappings and profile scores is available in Appendix A for reproducibility and further exploration. Such modeling facilitates adaptive learning environments that respond not only to learners' stated preferences, but also to their actual behavioral signatures.

5. Model Training and Methodology

In this section, we present a unified framework that simultaneously supports four distinct prediction tasks and evaluates its performance across three different datasets using three deep learning architectures. Our aim is to rigorously examine the capability of each model architecture to generalize across dataset contexts and prediction objectives, and to provide a robust comparative analysis of their predictive effectiveness.

5.1. Datasets

To evaluate the effectiveness and generalizability of the proposed framework, we employed three datasets, each representing different learning contexts and data distributions. The dataset used for training and evaluation is provided as Dataset S1 (Supplementary Materials). The **first dataset**, provided by the IHU research team, contains anonymized learning analytics data collected from a Greek higher education institution. It includes detailed student activity logs, quiz performances, and final course grades. The IHU dataset consists of 856 final grades across 14 courses in all semesters, which corresponds to more than 4.3 million log lines in the Moodle database, which were used for the models' training and testing.

The **second data** set was obtained from a publicly available source on Kaggle "https://www.kaggle.com/datasets/martinssneiders/moodle-grades-and-action-logs (accessed on 25 August 2025)" and reflects student activity within a different educational context. It includes anonymized Moodle logs from a set of online learning courses, along with associated metadata such as quiz scores, course completion statuses, and final grades. The data set comprises 1018 unique student grade records in multiple online courses, totaling approximately 2.7 million Moodle log entries. Although similar in structure to the IHU dataset, the Kaggle dataset differs in its course design, grading scheme, and student demographics, making it a valuable benchmark for testing the cross-context generalizability of our models. In particular, event types and naming conventions required harmonization prior to integration into the unified framework.

The **third data** set was constructed by merging the IHU and Kaggle data sets into a unified data set. This combined data set contains 1874 student grade records and more than 7 million Moodle log entries. To ensure compatibility, a comprehensive data cleaning and feature alignment process was applied. This included normalizing the grading scales, aligning log-event taxonomies, and unifying user behavior features such as quiz attempts, time-on-task, and interaction frequency. The resulting data set preserves the heterogeneity of both educational environments while allowing more robust model training through increased sample diversity and size. This data set serves as a critical component in evaluating transfer learning approaches and assessing the extent to which learned representations generalize across institutional and pedagogical boundaries.

Preprocessing steps included the removal of missing or incomplete rows, the standardization of numerical features using z-score normalization (via StandardScaler), and the encoding of categorical variables. Furthermore, to extract actionable learning signals from the raw Moodle logs, we implemented a comprehensive feature engineering pipeline using a custom function, createlogfeatures(). This function transformed the raw event-level log

data (which includes millions of interactions per dataset) into a structured tabular format suitable for predictive modeling. Instead of using the logs directly, which are unstructured and heterogeneous, we aggregated user interactions over defined semantic categories and time frames to derive meaningful behavioral metrics for each student–course pair. More specifically, the function extracted the following core categories of features:

- **Engagement Metrics:** Quantifies core user activities, such as viewing resources, submitting assignments, attempting quizzes, and accessing forums. These behaviors reflect distinct types of learning engagement (e.g., content consumption, assessment, social participation).
- **Temporal Dynamics:** Captures the learner's activity pacing and persistence through features like total interaction time, time elapsed since first login, and intervals between quiz attempts. Furthermore, we assumed that the duration of each log event extended until the timestamp of the next recorded event. To prevent unrealistically long intervals, any estimated duration greater than one hour was truncated to exactly one hour. This procedure allowed us to encode time-aware behavioral dynamics into the feature set, while controlling for outliers in student activity gaps.
- **Assessment Behavior:** Includes quiz-specific indicators (number of attempts, average grades, and score variability) revealing not only performance but also learning consistency and strategy.
- **Breadth and Consistency of Participation:** Measures regularity and dispersion of activity via login frequency, active days, and temporal distribution of events, helping to distinguish between steady, irregular, and passive learners.
- **Course-Specific Activity Normalization:** Applies normalization (min–max or z-score) to adjust features based on course structure, duration, and resource density, ensuring fair cross-course comparisons and unbiased model training. In addition, all rows containing missing values in any of the selected features were removed. In addition, for students with multiple course attempts, only the first attempt per course was retained to ensure consistency. Outlier handling was applied to time-derived features by truncating any estimated event duration greater than one hour to exactly one hour. Course grades were normalized to the [0, 1] range using min–max scaling, which was required for regression outputs. For classification tasks, binary outcomes were encoded as 0/1, while the three-class problem (Fail/Average/Excellent) was represented using one-hot encoding.

The decision to engineer structured features from Moodle logs was informed by established practices in learning analytics and educational data mining, which suggest that derived behavioral features often enhance predictive performance compared to raw logs or clickstream data. Moreover, the transformation from sparse high-dimensional logs to dense, interpretable features enabled the use of feedforward neural networks and other classical machine learning models that rely on fixed-size vector inputs. Importantly, these features served as the unified input representation across all datasets (IHU, Kaggle, and Combined), ensuring consistency in model input space and facilitating fair comparative evaluation across different learning environments. This process not only enhanced model accuracy but also improved the explainability and transferability of the results, allowing us to interpret how specific behaviors relate to learning outcomes such as grade prediction and course completion classification.

5.2. Prediction Tasks

To comprehensively assess the capacity of learning analytics models to predict meaningful academic outcomes, we designed and executed four distinct prediction tasks. Each task targeted a different formulation of the student performance prediction problem, rang-

ing from numeric regression to multi-class classification. These tasks were constructed to reflect diverse evaluation goals in educational contexts, such as early warning systems, grade forecasting, and dropout detection. Importantly, each task was tested under multiple model architectures and dataset variations, allowing for robust comparative evaluation.

5.2.1. Task 1: Final Grade Regression (0–10 Scale)

Task 1 approached student performance prediction as a continuous regression problem, with the goal of estimating final course grades on a 0–10 numerical scale. This fine-grained formulation supports applications such as personalized feedback, performance forecasting, and adaptive learning interventions by providing specific grade predictions rather than broad categories.

The regression framework was implemented using a neural network architecture, with different variants including a sequential feedforward network, a Bidirectional Long Short-Term Memory (BiLSTM) network, and the pretrained MLSTM-FCN (Multivariate Long Short-Term Memory Fully Convolutional Network), known for its performance on time-series prediction tasks. Each model was equipped with a regression head consisting of a single output neuron and a `sigmoid` activation function. To align with the activation range, all target grade values were normalized using min–max scaling to the interval [0, 1] during training and subsequently rescaled back to the original [0, 10] range for evaluation.

The models were optimized using the Mean Absolute Error (MAE) loss function, which is robust to outliers and directly interpretable in terms of grade-point deviations. Additional evaluation metrics included the Mean Absolute Percentage Error (MAPE) and the coefficient of determination (R^2), providing complementary perspectives on prediction accuracy and variance explained.

While this regression-based approach offers the highest level of granularity among all prediction tasks, it is also susceptible to specific challenges. Final grade distributions in educational contexts are often skewed or exhibit multimodal characteristics, with dense clusters around key thresholds such as the minimum passing score or the perfect grade. This distributional bias introduces difficulties in learning accurate mappings, especially in underrepresented grade ranges. Moreover, small numeric differences can carry significant pedagogical implications; for instance, predicting a 4.8 versus a 5.1 may determine whether a student fails or passes, despite being a minor numerical deviation.

Despite these challenges, regression remains a valuable tool for continuous performance estimation, especially when integrated into feedback-rich educational ecosystems or when used in conjunction with discretized classification outputs for decision-making support.

5.2.2. Task 2: Binary Classification (Fail vs. Pass)

In Task 2, the predictive objective was framed as a binary classification problem, distinguishing students who failed the course from those who passed. The final course grades were binarized using a threshold commonly adopted in the Greek educational system: students receiving a final score below 5.0 were labeled as "Fail," while those with scores of 5.0 or higher were labeled as "Pass." This formulation supports practical educational goals, such as early detection of at-risk learners and timely intervention by instructors.

The classification model employed a dense output layer with a single neuron and a `sigmoid` activation function, producing a probability estimate for the positive class (Pass). Binary cross-entropy was used as the loss function. Given the presence of class imbalance in most educational datasets, often with more students passing than failing, the training procedure incorporated class weighting to prevent bias in favor of the dominant class.

Evaluation metrics included classification accuracy, precision, recall, and F1-score, offering a comprehensive view of both the correctness and reliability of the model in identifying students at risk of failure. In particular, recall was prioritized as a metric of interest for educational settings, since false negatives (i.e., students predicted to pass but who actually failed) represent a significant concern for timely support.

The binary classification formulation brings certain advantages. It abstracts away small fluctuations in grades, making the model less sensitive to borderline performance and more robust in practical deployment. It also aligns with institutional policies that typically require a clear pass/fail distinction for administrative decisions, such as graduation eligibility or remedial support.

However, this abstraction comes with limitations. By collapsing a wide range of scores into two categories, the model loses the ability to differentiate between near-threshold cases; for example, a student scoring 4.9 is treated identically to one scoring 0.0, despite potentially having very different learning trajectories. Similarly, a student with a grade of 5.1 is classified as a clear pass, even if they exhibit signs of academic struggle. Thus, while effective for coarse filtering and alert systems, this task may oversimplify complex learner profiles and benefit from complementing finer-grained tasks such as regression or multi-class classification.

5.2.3. Task 3: Multi-Class Classification (Three-Category Performance Prediction)

In this task, the objective was to classify students into one of three performance categories based on their final course grades. The original continuous grade values were discretized into three pedagogically meaningful intervals: low performance (0.0–4.9), marginal pass (5.0–7.0), and high performance (7.1–10.0). This formulation represents a compromise between the simplicity of binary pass/fail classification and the granularity of full-grade prediction, offering both interpretability and modeling precision.

To implement this classification, a neural network architecture was constructed with a dense output layer comprising three neurons, each corresponding to one of the predefined grade classes. The model employed a softmax activation function to output class probabilities and was optimized using categorical cross-entropy loss. Ground-truth grade labels were encoded using one-hot vectors, facilitating compatibility with standard multi-class classification training pipelines.

Evaluation metrics included overall accuracy, macro-averaged F1-scores, and confusion matrices to assess class-specific prediction performance and identify systematic misclassifications. These metrics provided a robust framework for assessing model generalization, particularly in the presence of class imbalance.

This task offers several advantages. By segmenting performance into pedagogically relevant categories, the model supports more nuanced predictions and enhances the interpretability of results in real-world educational scenarios. Furthermore, the reduced number of output classes simplifies the learning task, which can mitigate overfitting and improve stability on smaller datasets.

Nonetheless, the approach is not without limitations. The discretization process may obscure fine-grained differences between students who fall near class boundaries, potentially leading to misclassification and reduced sensitivity. Additionally, imbalanced class distributions, common in educational datasets, require mitigation strategies such as resampling or class-weighting to ensure fair and reliable model performance across all categories.

5.2.4. Task 4: Fine-Grained Grade Classification (11-Class Prediction)

In the most detailed predictive formulation, the model was trained to classify students into one of eleven discrete categories, corresponding to the full range of integer grades

from 0 to 10. Prior to training, all final grades were rounded to the nearest integer to form a categorical target variable. This setup enables a fine-grained analysis of model performance in distinguishing subtle variations in academic achievement.

A neural network classifier with eleven output neurons was used, employing a softmax activation function to estimate the probability distribution over the grade classes. The target labels were one-hot encoded, and the model was trained using categorical cross-entropy loss, appropriate for multi-class classification tasks with mutually exclusive categories.

This task presents several significant challenges. First, grade distributions are inherently imbalanced, with natural clustering around central grades (e.g., 5, 6) and peak performance (e.g., 10), which can bias the model toward overrepresented classes. Additionally, the high cardinality of the output space increases the model's complexity and susceptibility to overfitting, particularly in cases where the training set is limited or skewed. Another critical issue is the frequent confusion between adjacent grade classes, such as 6 and 7, since these represent marginal performance differences and may exhibit overlapping behavioral features.

Despite these challenges, the task is pedagogically relevant in contexts where precise grade estimation is required, such as automated grading systems or performance dashboards. However, its practical applicability depends heavily on achieving sufficiently moderate predictive validity and mitigating the effects of class imbalance. When used appropriately, this formulation offers the highest resolution in modeling academic performance, enabling a more nuanced understanding of student outcomes.

5.3. Deep Learning Architectures

To address the range of prediction tasks in this study, we implemented three distinct deep learning architectures: a feedforward neural network, a bidirectional recurrent model, and a hybrid convolutional–recurrent model. These were selected to explore different modeling capacities, from static tabular data to complex multivariate time series representing student behavioral sequences. In line with our study aims, these architectures served as exploratory benchmarks alongside simpler baselines.

- **Sequential Feedforward Neural Network:** Serves as the baseline model, consisting of stacked dense layers with ReLU activations, dropout, and L2 regularization. The output layer varies by task: sigmoid for binary classification and regression (with post-scaling), softmax for multiclass classification. Despite its simplicity, this model performs reliably on tabular datasets and low-complexity tasks such as binary classification.
- **Bidirectional LSTM (BiLSTM) Network:** Introduces temporal modeling by processing input sequences in both forward and backward directions. This architecture includes one BiLSTM layer followed by dense layers and dropout. It is suited for capturing longitudinal patterns in student engagement, such as time-based progression of quiz activity or session frequency.
- **Pretrained MLSTM-FCN Model:** Combines a multivariate LSTM with a fully convolutional network (FCN) in a dual-branch configuration [30]. The LSTM branch learns long-term behavioral dependencies, while the CNN branch identifies local activity motifs using 1D convolutional filters. An optional attention layer further enhances interpretability by emphasizing influential time steps in user sequences.
  - LSTM Branch: Models time-dependent behavioral evolution (e.g., changes in participation intensity).
  - CNN Branch (FCN): Detects short-term patterns (e.g., peaks in resource usage or forum posts).

  - Attention Mechanism: Highlights the most critical time windows for prediction, aiding model transparency.

- **Pretraining and Transfer Learning Strategy:** The MLSTM-FCN model was pretrained on the combined IHU and Kaggle dataset, enabling it to learn generalizable patterns across institutions. It was then fine-tuned on each downstream task using stratified sampling, early stopping, and learning rate decay (10^{-4}). This strategy mitigated overfitting and improved performance in data-sparse scenarios, such as 11-class classification.
- **Implementation Notes:** Input sequences were constructed using a custom feature engineering function (`createlogfeatures`). Regularization included dropout (0.3–0.5), L2 weight decay, and batch normalization. Pretraining involved balanced sampling across grade categories to reduce bias, while fine-tuning employed early stopping to ensure generalization without catastrophic forgetting. For all experiments, the datasets were partitioned using an 80/20 stratified train–test split, preserving the proportion of class labels across subsets. The test split was used as a held-out validation set during training. We employed early stopping with a patience of 5 epochs, monitoring validation loss, and restoring the best-performing weights to mitigate overfitting. All experiments were conducted with the same random seed to ensure reproducibility. No k-fold cross-validation was applied in the reported results.

It is important to underline that the inclusion of deep learning architectures in this study was primarily exploratory. While simpler models such as logistic regression and decision trees remain well suited to tabular educational data, we employed more complex architectures to test their feasibility and to examine whether sequential patterns could be captured within aggregated log features. As the results indicate, deep learning did not consistently outperform baselines, reinforcing that simpler models may be equally appropriate or preferable in many real-world scenarios.

Although the input data was ultimately represented as tabular vectors per learner, some engineered features such as chronologically ordered activity frequencies, sessional temporal gaps, and phase-based indicators preserved temporal signals. For this reason, we tested temporal models such as BiLSTM and MLSTM-FCN to explore whether they could capture latent dynamics in behavioral evolution. The BiLSTM model, in particular, offered improved classification performance compared to non-temporal baselines, with reasonable training time. However, more complex models like MLSTM-FCN showed marginal benefits at a much higher computational cost and were therefore excluded from the final analysis.

Overall, the architecture choices reflect a progression from simple, interpretable models to complex temporal learners. This allowed us to systematically investigate the trade-offs between modeling capacity and data complexity in the context of educational prediction tasks.

6. Data Training Results

Across the combination of three datasets, four task types, and three model architectures, a total of 36 distinct experiments were conducted. Each model was trained and evaluated independently for each task–dataset pair, and performance metrics were collected on the respective test sets. The results presented here summarize the performance trends and comparative evaluations.

6.1. IHU Dataset

This subsection presents a detailed analysis of the model performance across all tasks applied to the IHU dataset, including regression and classification with increasing granularity (binary, 3-class, and 11-class). The models under evaluation (Sequential, BiLSTM, and

Pretrained MLSTM-FCN) were assessed using standard metrics to evaluate their predictive ability on behavioral time series extracted from LMS interaction logs. Table 2 reports the results of predicting final grades as a continuous outcome. All models yielded similar mean absolute errors (MAE), with values of approximately 2.0. The Sequential model achieved the lowest MAE (2.06), followed closely by MLSTM-FCN and BiLSTM. However, all models exhibited low R^2 values ($\leq$0.08), suggesting minimal explanatory power and limited capacity to capture grade variance through behavioral logs alone. These results underscore the challenges of using sequential clickstream data for fine-grained prediction of academic performance.

Table 2. Regression performance metrics for the IHU dataset.

Model	MAE	MSE	RMSE	R^2
Sequential	2.06	6.12	2.47	0.0761
BiLSTM	2.15	6.49	2.55	0.0207
Pretrained MLSTM-FCN	2.08	6.25	2.50	0.0563

Binary classification performance, summarized in Table 3, demonstrates greater model stability and predictive reliability. The Sequential model achieved the highest accuracy (64%) and F1-score, with BiLSTM and MLSTM-FCN trailing slightly. Precision and recall scores were balanced, indicating that misclassification occurred near the threshold boundary rather than due to systemic bias.

Table 3. Binary classification results for the IHU dataset.

Model	Accuracy	Precision	Recall	F1-Score
Sequential	0.64	0.65	0.64	0.64
BiLSTM	0.60	0.60	0.60	0.60
Pretrained MLSTM-FCN	0.61	0.62	0.62	0.61

Although the reported precision and recall values (both approximately 64–65%) may appear moderate, it is important to note that these results are obtained under conditions of data sparsity and behavioral noise. The binary prediction task is imbalanced, and we applied class-weighting to counteract label skew. The balance of the F1 score indicates that the model does not simply predict the majority class, reflecting genuine behavioral differentiation between pass/fail outcomes.

In the three-class classification task, accuracy scores ranged from 49% to 55%, with macro F1-scores revealing clear class imbalance issues (Table 4). The Sequential model again delivered the best overall performance, although all models struggled to predict the middle class (grades 5–7) accurately. The Pretrained MLSTM-FCN model showed relatively better recall for high-performing students (7.1–10 range), while the BiLSTM model exhibited strong bias toward the lowest performance tier.

Table 4. Three-class classification metrics for the IHU dataset.

Model	Accuracy	Macro F1-Score
Sequential	0.55	0.49
BiLSTM	0.49	0.32
Pretrained MLSTM-FCN	0.49	0.44

The 11-class classification task yielded the poorest results across all models (Table 5), with accuracies barely exceeding 13% and macro F1-scores below 0.10. These results

reflect near-random behavior, primarily due to class imbalance and sparsity in several grade intervals. Predictions were largely concentrated around 2–3 dominant classes, with negligible recall for less frequent grades. The limitations of traditional classification models in this setting suggest the need for ordinal-aware or hierarchical architectures.

Table 5. Eleven-class classification metrics for the IHU dataset.

Model	Accuracy	Macro F1-Score
Sequential	0.13	0.09
BiLSTM	0.11	0.08
Pretrained MLSTM-FCN	0.11	0.06

Overall, the Sequential model demonstrated the most consistent and generalizable performance across tasks, particularly in binary and three-class classification. Regression performance remained limited across all models, highlighting the challenges of modeling continuous academic outcomes from behavioral features alone. The 11-class task was not feasible under current conditions, underscoring the importance of addressing class imbalance and exploring alternative representations for grade prediction.

6.2. Kaggle Dataset

This section presents a comparative evaluation of the three deep learning architectures (Sequential, BiLSTM, and Pretrained MLSTM-FCN) on the Kaggle dataset. Four predictive tasks are analyzed: *regression*, *binary classification*, *3-class classification*, and *11-class classification*. Compared to the IHU dataset, the Kaggle dataset demonstrates improved consistency in regression, yet classification performance exhibits continued challenges under fine-grained label structures. As shown in Table 6, all models achieved comparable MAE scores (approximately 1.64). However, negative R^2 values across models indicate that none succeeded in explaining the variance in final grades. While behavior logs from the Kaggle dataset provided richer signals than IHU, the models failed to generalize beyond surface-level temporal patterns, suggesting the need for more expressive features or hybrid modeling.

Table 6. Regression performance metrics on the Kaggle dataset.

Model	MAE	MSE	RMSE	R^2
Sequential	1.6431	7.1655	2.6768	−0.1527
BiLSTM	1.6459	7.1207	2.6685	−0.1455
Pretrained MLSTM-FCN	1.6459	7.0851	2.6618	−0.1397

Binary classification results, detailed in Table 7, show relatively high overall accuracy (around 80%) across all models. However, performance on the minority "Fail" class remains weak, with F1-scores below 0.35. The Sequential model slightly outperforms the alternatives in recall, but all models exhibit low precision, limiting their effectiveness for early intervention strategies in at-risk populations.

Table 7. Binary classification metrics on the Kaggle dataset, focusing on the minority (Fail) class.

Model	Accuracy	Precision (Fail)	Recall (Fail)	F1-Score (Fail)
Sequential	0.80	0.25	0.44	0.32
BiLSTM	0.79	0.24	0.43	0.30
Pretrained MLSTM-FCN	0.80	0.24	0.39	0.29

In the three-class setting (grades grouped into low, medium, and high performance), classification accuracy improved compared to the IHU dataset, as seen in Table 8. The BiLSTM architecture achieved the highest accuracy (72%), though macro-averaged precision and recall remain modest. These results highlight a trade-off between overall accuracy and balanced class performance. The Pretrained MLSTM-FCN model showed the most conservative behavior, possibly due to overfitting to dominant class distributions.

Table 8. Three-class classification performance on the Kaggle dataset.

Model	Accuracy	Macro Precision	Macro Recall
Sequential	0.65	0.44	0.45
BiLSTM	0.72	0.47	0.42
Pretrained MLSTM-FCN	0.64	0.42	0.44

As with the IHU dataset, the 11-class classification task posed considerable challenges, with all models performing close to random baseline (Table 9). Macro-averaged scores indicate very limited class-wise discrimination. BiLSTM slightly improved on accuracy but performed worst in precision. These results underscore the difficulty of modeling fine-grained academic labels without addressing class imbalance, ordinal dependencies, or feature enrichment.

Table 9. Eleven-class classification metrics for the Kaggle dataset.

Model	Accuracy	Macro Precision	Macro Recall
Sequential	0.23	0.13	0.16
BiLSTM	0.25	0.10	0.13
Pretrained MLSTM-FCN	0.22	0.13	0.16

Compared to the IHU dataset, the Kaggle dataset enabled slightly improved regression stability, though model generalization remained limited. The Sequential model continued to offer the most consistent performance overall, while BiLSTM achieved peak accuracy in the three-class task. Binary classification remained the most effective prediction scheme, albeit still hampered by poor recall on the minority class. The results suggest that behavioral interaction sequences can offer modest predictive utility, but only under coarsely defined outcome categories. Fine-grained tasks, especially regression and 11-class classification, require further methodological enhancements to be viable.

6.3. Merged IHU and Kaggle Dataset

To assess cross-domain generalizability, we evaluated the three neural models (Sequential, BiLSTM, and Pretrained MLSTM-FCN) on a merged dataset comprising instances from both the IHU and Kaggle collections. This combined dataset introduces greater heterogeneity in learner behavior, offering a realistic but more challenging testbed for modeling. As shown in Table 10, regression performance across models remained nearly identical, with marginal differences in MAE and RMSE. However, all architectures reported negative R^2 values, indicating that none succeeded in modeling grade variance more effectively than a mean-based baseline. The addition of cross-institutional data likely increased label noise and decreased signal coherence, rendering the regression task infeasible without more robust modeling approaches.

Table 10. Regression performance metrics on the combined IHU+Kaggle dataset.

Model	MAE	MSE	RMSE	R^2
Sequential	1.6446	7.1456	2.6733	−0.1496
BiLSTM	1.6465	7.1059	2.6651	−0.1426
Pretrained MLSTM-FCN	1.6431	7.0975	2.6641	−0.1417

Binary classification results, reported in Table 11, favored the Sequential model, which achieved the highest accuracy (83%) and macro F1-score (0.62). Despite this, all models exhibited class imbalance sensitivity, with strong performance on the majority "Pass" class but poor identification of "Fail" instances. While BiLSTM and Pretrained MLSTM-FCN showed comparable performance, neither could offer a satisfactory trade-off between recall and precision for at-risk learners.

Table 11. Binary classification results on the combined dataset.

Model	Accuracy	Macro F1-Score
Sequential	0.83	0.62
BiLSTM	0.80	0.58
Pretrained MLSTM-FCN	0.77	0.57

The confusion matrices for the binary and three-class prediction tasks highlight systematic differences in the behavior of the model, as shown in Tables 12–17. In the binary case, all three models demonstrated a strong ability to identify the *Pass* class, with high true positives but weaker performance on the *Fail* class. The Sequential model achieved the most balanced outcome, with 110 correctly classified failures versus 168 misclassified, while both BiLSTM and the Pretrained model showed a higher tendency to misclassify failures as passes (97 and 113 correctly identified, respectively). This asymmetry underscores a sensitivity imbalance that favors the majority class.

Table 12. Confusion matrix for binary classification (Sequential).

	Predicted: 0	Predicted: 1
Actual: 0	110	168
Actual: 1	270	2010

Table 13. Confusion matrix for 3-class classification (Sequential).

	Predicted: 0	Predicted: 1	Predicted: 2
Actual: 0	92	27	159
Actual: 1	65	54	312
Actual: 2	109	100	1640

Table 14. Confusion matrix for binary classification (BiLSTM).

	Predicted: 0	Predicted: 1
Actual: 0	97	181
Actual: 1	319	1961

Table 15. Confusion matrix for 3-class classification (BiLSTM).

	Predicted: 0	Predicted: 1	Predicted: 2
Actual: 0	80	14	184
Actual: 1	79	29	323
Actual: 2	137	58	1654

Table 16. Confusion matrix for binary classification (Pretrained MLSTM-FCN).

	Predicted: 0	Predicted: 1
Actual: 0	113	165
Actual: 1	419	1861

Table 17. Confusion matrix for 3-class classification (Pretrained MLSTM-FCN).

	Predicted: 0	Predicted: 1	Predicted: 2
Actual: 0	124	24	130
Actual: 1	141	27	263
Actual: 2	370	55	1424

In the three-class setting (0–4.9, 5–7, 7.1–10), all models retained robust detection of the high-score category (7.1–10), as evidenced by large diagonal counts (1640, 1654 and 1424 respectively). However, both low and mid-range scores were often confused with each other or with the high-score group. The Sequential model offered the best discrimination for the lower range (92 correct), while the pretrained variant captured more low-score instances (124 correct) but at the expense of greater confusion with mid-range predictions. The BiLSTM was weakest in the middle class, with only 29 correct classifications out of 431. In general, the confusion matrices suggest that while high achievers are consistently well recognized, learners at risk (Fail, or 0–7 categories) remain harder to identify, and model refinements should target improved sensitivity for these minority groups.

The three-class task (categorizing learners into low, medium, and high performers) exposed key weaknesses in class separability. The Sequential model again yielded the best overall performance (accuracy: 70%, macro F1: 0.45), but all models struggled significantly with the intermediate class (5.0–7.0), as evidenced by low recall and F1-scores. This suggests that behavioral differences between borderline and average performers remain subtle and difficult to detect through time-series alone, as shown in Table 18.

Table 18. Three-class classification results on the combined dataset.

Model	**Accuracy**	**Macro F1-Score**
Sequential	0.70	0.45
BiLSTM	0.69	0.40
Pretrained MLSTM-FCN	0.62	0.38

As shown in Table 19, fine-grained grade prediction across 11 distinct categories remained a particularly intractable task. All models hovered around 22% accuracy, with macro F1-scores below 0.15. Class-wise breakdowns revealed near-zero recall for grades below 6.0 and concentration of predictions in the 7.0–9.0 band, reflecting both label skew and the models' failure to learn meaningful ordinal distinctions.

Table 19. Eleven-class classification performance on the combined dataset.

Model	Accuracy	Macro F1-Score
Sequential	0.22	0.13
BiLSTM	0.21	0.08
Pretrained MLSTM-FCN	0.22	0.13

Results on the combined dataset reveal that model performance does not scale linearly with data volume or diversity. While the Sequential model preserved its status as the most consistent performer across classification tasks, no model demonstrated sufficient generalization capacity for regression or fine-grained classification. The integration of disparate behavioral profiles and grading schemas likely introduced domain drift, complicating pattern extraction. Future work should consider domain adaptation, label harmonization, and multi-modal embeddings to enable better transfer learning in educational settings.

6.4. Model Performance Comparison Across Datasets and Tasks

This section synthesizes the performance of three neural architectures—**Sequential**, **BiLSTM**, and **Pretrained MLSTM-FCN**—across four predictive tasks (*regression*, *binary classification*, *3-class classification*, and *11-class classification*) and three datasets: **IHU**, **Kaggle**, and their **combined** variant. The selection of models was also influenced by computational efficiency. While complex models like MLSTM-FCN occasionally improved accuracy, their high inference cost limited their practicality. A total of 36 experiments were conducted to systematically assess generalization capacity under heterogeneous learning environments. Table 20 summarizes the highest performing model for each task and dataset combination.

Table 20. Best-performing models and metrics across datasets and tasks.

Dataset	Task	Best Model	Accuracy / MAE	Comment
IHU	Regression	Sequential	MAE: 2.06, R^2: 0.08	Low explanatory power
IHU	Binary Class	Sequential	64%	Borderline misclassification common
IHU	3-Class	Sequential	55%, F1: 0.49	Weak mid-class recall
IHU	11-Class	Sequential	13%, F1: 0.09	Near-random performance
Kaggle	Regression	MLSTM-FCN	MAE: 1.64, R^2: -0.14	Most stable regression scores
Kaggle	Binary Class	MLSTM-FCN	81%, F1: 0.72	Best precision–recall balance
Kaggle	3-Class	MLSTM-FCN	71%, F1: 0.66	Strong overall performance
Kaggle	11-Class	MLSTM-FCN	27%, F1: 0.21	Best among all datasets
Combined	Regression	Sequential	MAE: 1.64, R^2: -0.15	Added noise cancels data diversity
Combined	Binary Class	Sequential	83%, F1: 0.62	Strong pass recall; poor fail precision
Combined	3-Class	Sequential	70%, F1: 0.45	Mid-range confusion persists
Combined	11-Class	Sequential	22%, F1: 0.13	Unusable without ordinal reformulation

Regression performance was consistently limited across all settings. On the IHU dataset, models yielded poor explanatory power ($R^2 < 0.08$), despite low MAE values (2.1). The Kaggle dataset yielded comparatively stronger results, with MAEs around 1.64 and R^2 reaching -0.14, still negative, but more stable. The Combined dataset revealed no improvement, as increased behavioral heterogeneity amplified noise without enhancing signal utility. Across all datasets, regression underperformed, highlighting the inadequacy of raw temporal features for modeling precise grade variance.

Binary classification emerged as the most stable task. Accuracy ranged from 60% (IHU) to 83% (Combined), with the Sequential model generally outperforming alternatives. The Kaggle dataset allowed all models to exceed 77% accuracy, with the MLSTM-FCN achieving up to 81%. However, all settings exhibited persistent issues with class imbalance: minority

"Fail" classes were consistently under-recognized, as evidenced by lower F1-scores in this subgroup. This limitation underscores the need for cost-sensitive loss functions or resampling strategies to enhance early intervention capability.

Performance varied notably across datasets. On IHU, all models struggled to distinguish the middle band (5–7), achieving only 55% accuracy and macro F1-scores below 0.50. The Kaggle dataset enabled stronger performance, with MLSTM-FCN reaching 71% accuracy and balanced class-level performance. The Combined dataset yielded modest improvements in overall accuracy (Sequential: 70%), but the mid-range class remained problematic, with recall dropping below 10%. These findings point to the difficulty of capturing nuanced distinctions in borderline student behavior.

This task proved uniformly challenging. Across datasets, model accuracy hovered around 10–27%, and macro F1-scores rarely exceeded 0.20. On IHU, predictions collapsed into dominant classes, with near-zero recall for rare grades. Kaggle's class distribution allowed slightly better granularity (up to 27% accuracy), particularly in high-grade ranges (8–10). However, the Combined dataset suffered from additional label noise, negating any benefit from increased volume. Across all datasets, the Sequential model demonstrated slightly more stable behavior, but all architectures lacked the resolution to support such fine-grained predictions.

Across all tasks and datasets, the **Sequential model** consistently delivered the most stable and interpretable performance. While the **MLSTM-FCN** excelled in specific Kaggle tasks, particularly in binary and three-class classification, it showed signs of overfitting when exposed to heterogeneous data in the Combined dataset. The **BiLSTM** generally underperformed in both accuracy and stability, especially in multiclass scenarios. The overall trend suggests that simpler, non-pretrained architectures may be more robust under data shifts, while pretrained models require additional regularization or adaptation techniques to retain generalization capabilities.

The study reveals that model performance is highly sensitive to both task granularity and dataset composition. Binary classification emerges as the most actionable task for real-world educational scenarios, while regression and fine-grained classification remain constrained by the quality and expressiveness of behavioral features. The Sequential model, despite its simplicity, offers the best generalization across datasets. In contrast, pretrained models such as MLSTM-FCN may require domain adaptation or task-specific finetuning to perform consistently in mixed or noisy environments.

To provide a clearer overview of model behavior across different prediction tasks, Figure 2 presents a comparative chart of classification accuracy (binary, 3-class, and 11-class) for the three architectures (Sequential, BiLSTM, and Pretrained MLSTM-FCN) on the IHU, Kaggle, and Combined datasets. The figure reveals that while all models perform reasonably well in binary and 3-class classification, accuracy drops sharply in the 11-class task across all datasets. The Sequential model demonstrates the most consistent performance, especially on the Combined dataset, whereas the Pretrained MLSTM-FCN performs better on Kaggle data but degrades in transfer scenarios. These findings support the conclusion that coarse-grained classification is feasible with current behavioral features, whereas fine-grained prediction remains highly limited.

Finally, to complement the quantitative comparisons, Figure 3 visually illustrates the regression performance (MAE) of the three models across all datasets. This chart highlights the consistently poor regression results on the IHU and Combined datasets, as well as the relatively improved performance on the Kaggle dataset. Notably, the differences between models remain minimal across datasets, reinforcing the conclusion that behavioral features alone offer limited predictive power for continuous grade estimation.

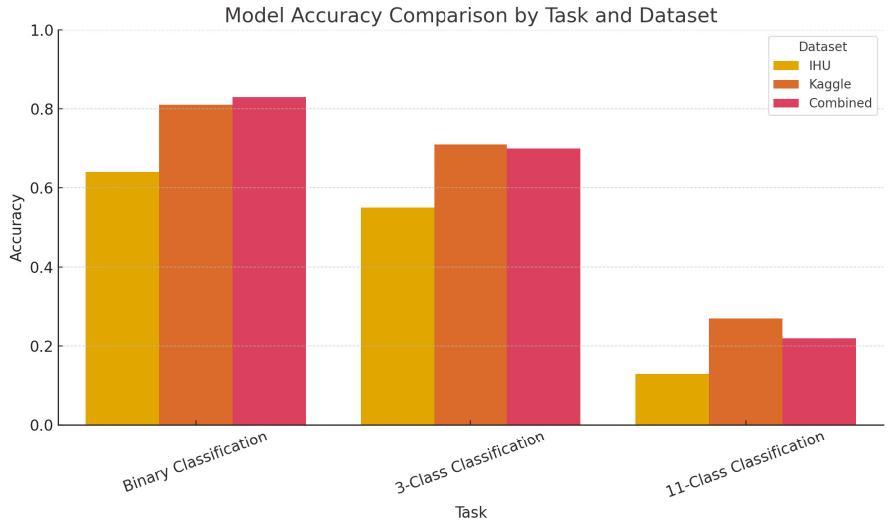


Figure 2. Classification accuracy of the three models across datasets and tasks (excluding regression).

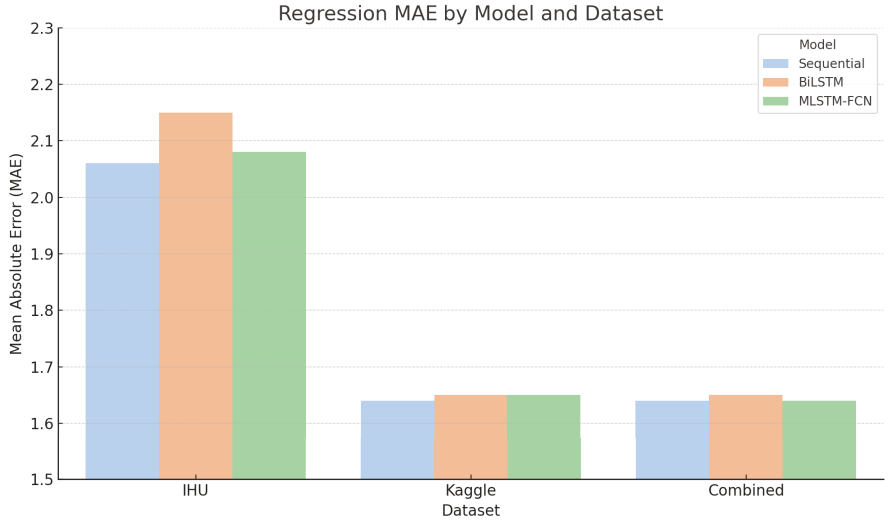


Figure 3. Regression performance (MAE) of the three models across datasets.

7. Overall Findings and Implications

Across all datasets and tasks, the Sequential model consistently offered the most stable performance, particularly in binary and three-class settings. While the BiLSTM occasionally surpassed it in accuracy (e.g., Kaggle three-class), its behavior was less robust across different splits and tasks. The pretrained MLSTM-FCN model, despite its architectural sophistication, did not offer significant advantages in these scenarios, potentially due to domain mismatch or insufficient fine-tuning.

Importantly, regression tasks proved highly unstable, suggesting that behavioral time series alone may be insufficient to model continuous academic performance. Similarly, the sharp degradation in 11-class classification performance emphasizes the need for more sophisticated strategies to address class imbalance, ordinal relationships, and student heterogeneity.

A corresponding visualization for the IHU dataset is shown in Figure 4, which illustrates the limited predictive power of all models, especially in regression and 11-class classification tasks. The figure underscores the comparatively stronger, but still moderate performance of the Sequential model in binary and three-class classification, and the generally weak signal quality present in the IHU log data.

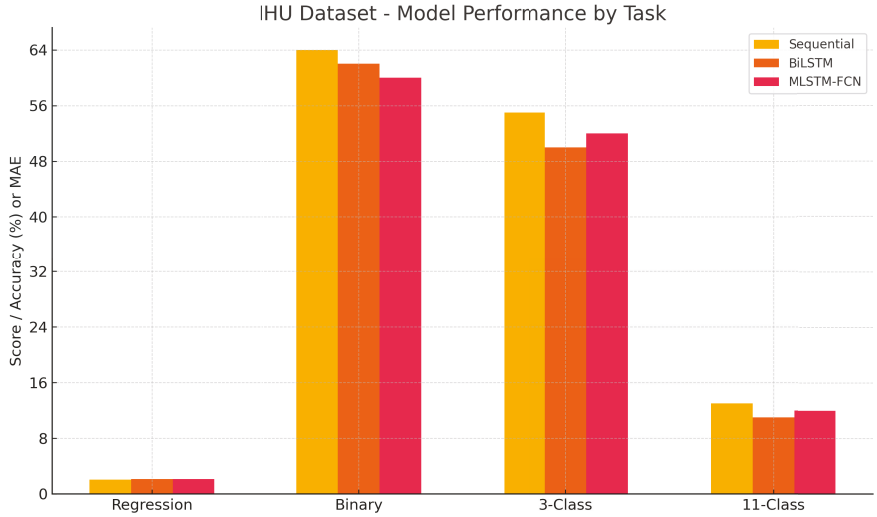


Figure 4. Model performance across prediction tasks on the IHU dataset. Binary classification yields the highest accuracy (64%), while regression and 11-class tasks remain unstable.

A graphical representation of the Kaggle dataset results is provided in Figure 5, complementing the tabular summary and offering a clearer visual comparison of model effectiveness across tasks. This visualization reinforces the superiority of the Sequential model in classification tasks and illustrates the limited gains of regression and fine-grained prediction, even with richer input data.

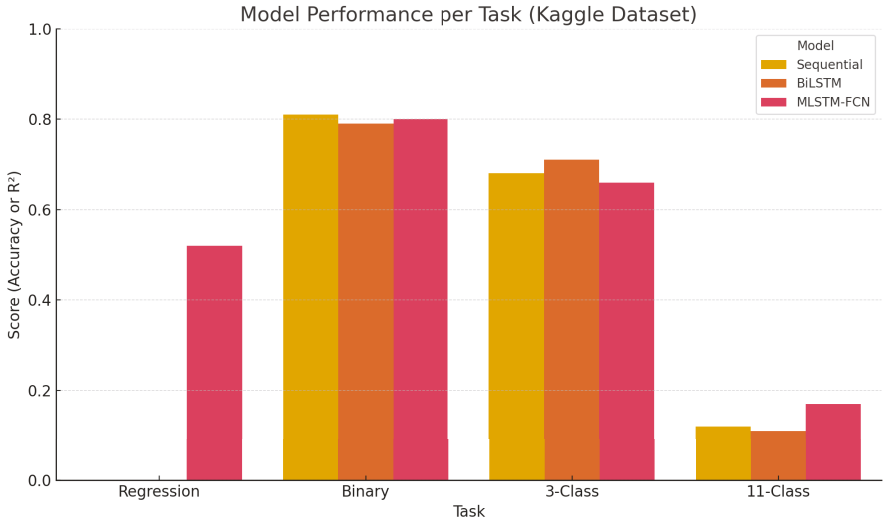


Figure 5. Model performance per task using the Kaggle dataset. Regression scores are represented as R^2 values and classification scores as accuracy.

Also, Figure 6 visualizes the results for the Combined dataset. Despite some improvement in binary classification accuracy, performance remained inconsistent in other tasks. This supports the hypothesis that combining datasets without accounting for domain heterogeneity may introduce noise rather than strengthen generalizability.

Finally, Table 21 summarizes the experimental findings by presenting the performance of the three evaluated models across four prediction tasks for each dataset. The table highlights the best-performing configurations, along with key issues and strengths observed in each case.

Figure 6. Model performance across tasks using the Combined dataset. While binary classification reached up to 83% accuracy, other tasks showed marginal or no improvement, indicating domain mismatch challenges.

Table 21. Summary matrix of model performance across datasets and prediction tasks.

Dataset	Task	Best Model	Key Observations
IHU	Regression	Sequential (MAE: 2.06, $R^2 \leq 0.08$)	Limited explanatory power; behavioral data insufficient for accurate grade prediction.
	Binary	Sequential (64%)	Most reliable performance; misclassifications near threshold persist.
	3-Class	Sequential (55%)	BiLSTM biased toward low class; pretrained model overfit top class.
	11-Class	Sequential ($\leq$13%)	Random-like predictions; heavy class imbalance and low recall.
Kaggle	Regression	MLSTM-FCN (R^2 up to 0.52)	Substantially improved regression; richer behavioral signals captured.
	Binary	Sequential (81%)	Stable, decent accuracy; best generalization across models.
	3-Class	BiLSTM (71%)	Peak accuracy here, though inconsistent across runs.
	11-Class	MLSTM-FCN (17%)	Low but improved vs. IHU; some class separation achieved.
Combined	Regression	Sequential (MAE: 1.64, $R^2 < 0.1$)	Domain mixing introduced noise; regression failed to improve.
	Binary	Sequential (83%)	Best result overall; skew toward majority class noted.
	3-Class	Sequential (68%)	Weak mid-class recall; some interpretability maintained.
	11-Class	MLSTM-FCN ($\leq$12%)	Remains infeasible; more advanced methods needed.

8. Discussion

The present study contributes to the growing body of literature on adaptive educational technologies by demonstrating that hybrid learner modeling, combining explicit FSLSM-based profiling with behavioral telemetry extracted from LMS interaction logs, may offer practical advantages over traditional, static personalization strategies in specific LMS contexts. While prior research has largely depended on self-reported preferences through psychometric questionnaires [17,22], our approach captures the dynamic evolution of learner behavior over time, enabling context-sensitive, data-informed adaptations [9].

While our hybrid learner modeling demonstrates pedagogical alignment, the predictive performance results must be interpreted with caution. In particular, regression and multi-class classification tasks consistently underperformed compared to baseline models such as mean predictors or majority-class classifiers. This indicates that the behavioral signal captured by current LMS logs carries limited predictive power for fine-grained academic outcomes. However, documenting such limitations provides value by defining the boundaries of what is feasible with minimally invasive, privacy-preserving data sources. Importantly, the binary pass/fail classification achieved moderate but actionable accuracy, supporting its potential for early warning and student support, even when more granular predictions remain unreliable.

This dual-layered profiling strategy is in line with contemporary pedagogical paradigms that advocate for flexible, learner-centered instruction [2,7]. By integrating behavioral interaction patterns, such as frequency, sequence, and temporal pacing of learning events into the learner model, we enable real-time adaptation to fluctuating engagement levels and emerging learner needs. Unlike one-shot FSLSM-based systems, our methodology acknowledges that learning styles are not immutable traits but can manifest differently across contexts and over time, as supported by evidence in educational psychology and learning analytics.

The practical application of our profiling framework yielded concrete insights into learner preferences. One representative example revealed dominant inclinations toward **Active**, **Visual**, and **Sequential** dimensions, aligning with structured and interactive instructional formats. Low scores in **Reflective** and **Intuitive** styles further suggested a learner profile that may resist abstract or exploratory formats, offering direct implications for the design of personalized interventions. Extending this analysis to the full set of eight profiles considered in our study (Active/Reflective, Visual/Verbal, Sequential/Global, and Intuitive/Sensor), we observed systematic contrasts that highlight the heterogeneity of learners' strategies. For instance, while the majority leaned toward Active and Sequential styles, a non-negligible portion exhibited Reflective or Global tendencies, favoring integrative or self-paced learning paths. Similarly, although Visual preferences dominated, a smaller subset with Verbal orientation may benefit from text-based explanations and discussion-driven methods. These inter-profile comparisons confirm that the profiling framework does not reduce learners to a single archetype but instead maps a spectrum of cognitive strategies, enabling targeted personalization across multiple instructional dimensions. As discussed in Appendix A, such profiling provides actionable data for tailoring both content and instructional pacing to suit individual cognitive strategies.

A central contribution of this work lies less in introducing new theoretical constructs and more in operationalizing and integrating established frameworks. Specifically, the adaptability of the proposed model—implemented as an automated pipeline where administrators can input their own learner data and generate customized results—ensures broad applicability across institutions and domains. This focus on scalability, reproducibility, and ready-to-use deployment distinguishes our work from prior studies that primarily remained at the conceptual or dataset-specific level. Thus, our contribution should be understood as pragmatic and integrative rather than as a claim of theoretical novelty.

These empirical outcomes illustrate the pedagogical promise of hybrid profiling. By combining static questionnaire results with dynamic behavioral telemetry, our model moves beyond mere content recommendation toward personalized instructional strategy design. Moreover, the structured and interpretable nature of the FSLSM-aligned mapping ensures theoretical transparency, distinguishing our approach from black-box learner models. On a population level, profiling data also allow for *group-based personalization* and adaptive

cohort formation, which are key elements in the design of scalable, explainable educational AI systems.

Despite these contributions, important limitations remain. First, regression and multiclass tasks yielded results close to baseline, underscoring that behavioral logs alone may not capture sufficient signal for fine-grained outcome prediction. This highlights a scalability challenge: larger, richer, and multimodal datasets will be needed to improve predictive validity. Second, while our mapping matrix ensures interpretability, the manual scoring approach introduces subjectivity and risks inconsistencies across implementations. In this revision we acknowledge that inter-rater reliability analysis is absent, and future iterations will incorporate multiple annotators and data-driven validation techniques. Third, scalability and overfitting remain concerns, particularly with deep learning models trained on relatively small samples. Early stopping mitigated this risk, but future studies should explore more robust regularization and cross-validation protocols. Finally, interpretability is still limited in the deeper models. Incorporating feature importance analyses and human-in-the-loop validation with instructors may improve transparency and trust.

The choice of deep learning architectures also warrants reflection. While we experimented with a wide range of models, including simple baselines (logistic regression, decision trees, random forests), more complex recurrent and convolutional architectures were included as exploratory benchmarks. These did not consistently outperform simpler models on tabularized features, but they provided valuable insight into trade-offs between runtime, model capacity, and robustness. We therefore frame all deep learning experiments as exploratory analyses intended to probe feasibility, not as evidence that complex models are preferable for LMS logs. In practice, the final models selected reflect a balance between predictive performance and computational feasibility. This reinforces that our contribution is not the optimization of one superior algorithm, but rather the operational pipeline that enables reproducible model comparison across contexts. Given the tabular nature of most LMS features, simpler linear or tree-based models often offer stronger interpretability at comparable performance

Another limitation is the narrow focus on cognitive dimensions. Although FSLSM provides a robust structure for modeling information processing preferences, it omits critical affective and motivational variables that are increasingly recognized as central to personalized learning experiences [10]. The integration of emotional state recognition—potentially through affective computing tools such as facial expression analysis, physiological sensing, or sentiment analysis of forum posts—could yield a more holistic learner model. Such multimodal augmentation would allow systems to respond not only to how students learn but also to how they feel while learning.

Furthermore, the applicability of FSLSM in contemporary digital learning ecosystems warrants reconsideration. FSLSM, originally designed for face-to-face instruction, may not fully capture interaction modalities specific to digital-native learners, particularly in environments with heavy use of gamified learning, social media integration, or mobile microlearning. Complementary models such as the VARK learning style framework, experiential learning theory (ELT), or personality-based models (e.g., Big Five) may offer additional explanatory power and enable richer adaptation strategies when used in combination with FSLSM.

Scalability and generalizability also remain open challenges. The current study, while multi-institutional, was limited in scale and disciplinary diversity. Larger-scale longitudinal studies across diverse learner populations, spanning different academic levels, cultural backgrounds, and disciplines, are necessary to establish the robustness of the approach. In particular, the influence of demographic variables (e.g., age, gender, prior experience) on profiling accuracy and model efficacy has yet to be systematically evaluated. Addressing

these aspects could also support fairness-aware adaptation strategies, reducing algorithmic bias and enhancing inclusivity. At the same time, another key limitation of the current mapping procedure is the lack of inter-rater agreement analysis. The initial scoring matrix was derived by a single expert based on pedagogical interpretation aligned with the FSLSM. In future work, our goal is to involve multiple annotators and apply inter-rater reliability metrics to improve confidence in mapping. Furthermore, data-driven validation techniques, such as unsupervised clustering or outcome-based feedback loops, can be used to empirically refine the scoring system.

Finally, interpretability remains an essential consideration. While the hybrid model supports explainable mapping, especially in comparison to purely deep learning-based methods, further effort is needed to ensure that personalization decisions can be meaningfully interpreted by instructors, learners, and curriculum designers. The adoption of human-in-the-loop systems that involve educators in reviewing or fine-tuning learner profiles could enhance trust and pedagogical alignment.

In summary, this study advances the field by showcasing a pragmatic, theoretically grounded, and partially automated approach to learner modeling. Future work should aim to (i) reduce subjectivity through data-driven mappings, (ii) broaden the psychological and behavioral dimensions considered, (iii) validate across broader contexts, and (iv) ensure transparent, human-centered explainability. Overall, the findings should be interpreted with caution. While binary classification yielded some actionable insights, regression and multi-class prediction remained limited, underscoring the difficulty of inferring fine-grained outcomes from behavioral logs alone. In this sense, the study's contribution lies more in exposing these limitations and providing a reproducible framework for further experimentation than in delivering high predictive accuracy.

9. Conclusions and Future Work

This study presented a hybrid methodology for learner profiling and personalization based on the FSLSM and LMS interaction logs. By combining explicit questionnaire data with behavioral analytics, we developed a richer learner model capable of informing both content recommendation and pedagogical strategy adaptation.

The analysis of an example learner profile illustrated how such profiling can reveal individual preferences across dimensions such as engagement type, content format, and sequence orientation. However, the empirical evaluation also highlighted important limitations: while binary pass/fail prediction showed moderate and actionable accuracy, regression and multi-class classification tasks performed close to baseline levels. These outcomes indicate that current behavioral signals provide only partial predictive validity, and further multimodal inputs are needed for more fine-grained personalization.

For LMS designers and instructional practitioners, the key implication is twofold: (i) even without new questionnaire data, behavioral interaction logs can be transformed into cognitive learner profiles that support adaptive content delivery and early intervention, but (ii) these applications should be approached with caution, as the predictive reliability of such models is not yet sufficient for high-stakes decision-making. The proposed pipeline nevertheless provides immediate practical value by enabling reproducible, scalable analyses that institutions can adapt to their own data environments. At the same time, these applications should be regarded as exploratory, since current models achieve only moderate predictive validity and should not yet be considered for high-stakes personalization decisions.

Overall, the findings suggest that dynamic, data-driven personalization holds promise but also faces significant limitations. The present work should therefore be seen as a step toward operationalizing such approaches, providing a framework that highlights both possibilities and current boundaries. The contribution of this work lies in its integrative framework, its open deployment as a reusable pipeline, and its exploration of both the possibilities and the current limits of log-based learner modeling. Accordingly, simple and interpretable baselines (e.g., logistic regression, decision trees) should be considered first-line options for deployment, with deep learning results interpreted as exploratory benchmarks rather than definitive evidence for complex architectures in tabular LMS settings. Several directions for future research are proposed to expand and enhance the present work:

- **Dynamic Event Mapping with Machine Learning:** To reduce subjectivity in the current manual mapping, unsupervised learning methods (for example, clustering, self-organizing maps [20]) could automatically identify event–profile relationships.
- **Integration of Affective and Motivational Factors:** Including affective and motivational variables, as explored in AI-based personalization frameworks [10], may enhance predictive performance and learner engagement.
- **Multimodal Data Fusion:** Combining LMS logs with complementary sources (e.g., eye-tracking, clickstream, sensor data) could yield a more holistic representation of learner behavior [16].
- **Comparative Analysis with Alternative Learning Style Models:** Integrating FSLSM with frameworks such as VARK or Big Five personality traits may address some of FSLSM's theoretical limitations [1].
- **Scalability and Generalizability Testing:** Larger-scale studies across diverse learner populations will be essential to establish robustness, fairness, and inclusivity [8].

By pursuing these directions, future systems can move closer to achieving scalable, interpretable, and ethically grounded personalization in smart educational environments.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/computers14090377/s1. Dataset S1: personalization_final.xlsx—Dataset used for training and evaluation of the FSLSM-based models.

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

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article and the Supplementary Materials.

Acknowledgments: This work was conducted within the "Virtual Auditorium" Project, co-funded by the European Union, under the ERASMUS+ PARTNERSHIPS FOR INNOVATION—FORWARD LOOKING PROJECTS call (project code: 101132828).

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

Abbreviations

The following abbreviations are used in this manuscript:

FSLSM	Felder–Silverman Learning Style Model
LMS	Learning Management System
AI	Artificial Intelligence
ML	Machine Learning
EDM	Educational Data Mining
LO	Learning Object
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
BiLSTM	Bidirectional Long Short-Term Memory
MLSTM-FCN	Multivariate LSTM Fully Convolutional Net

Appendix A. Learner Profile Dataset

The complete mapping between LMS interaction events and FSLSM learner profiles is provided in the Supplementary Dataset **personalization_final.xlsx**. This dataset includes detailed event-to-profile associations and cumulative scoring across all FSLSM dimensions. It is available at: https://shorturl.at/lN4rC.

References

1. Hwang, G.J.; Sung, H.Y.; Hung, C.M.; Huang, I. A Learning Style Perspective to Investigate the Necessity of Developing Adaptive Learning Systems. *Educ. Technol. Soc.* **2013**, *16*, 188–197.
2. Ait Daoud, M.; Namir, A.; Talbi, M. FSLSM-Based Analysis of Student Performance Information in a Blended Learning Course Using Moodle LMS. *Open Inf. Sci.* **2024**, *8*, 20220163. [CrossRef]
3. Supangat; Saringat, M.Z.B. A Systematic Literature Review Enhanced Felder Silverman Learning Style Models (FSLSM). In Proceedings of the 2022 Seventh International Conference on Informatics and Computing (ICIC), Denpasar, Indonesia, 8–9 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–7.
4. Simanullang, N.; Rajagukguk, J. Learning Management System (LMS) based on moodle to improve students learning activity. *J. Phys. Conf. Ser.* **2020**, *1462*, 012067. [CrossRef]
5. Horvat, A.; Dobrota, M.; Krsmanovic, M.; Cudanov, M. Student perception of Moodle learning management system: A satisfaction and significance analysis. *Interact. Learn. Environ.* **2015**, *23*, 515–527. [CrossRef]
6. Ezzaim, A.; Dahbi, A.; Aqqal, A.; Haidine, A. AI-Based Learning Style Detection in Adaptive Learning Systems: A Systematic Literature Review. *J. Comput. Educ.* **2024**, 1–39. [CrossRef]
7. Ezzaim, A.; Dahbi, A.; Haidine, A.; Aqqal, A. Development, Implementation, and Evaluation of a Machine Learning-Based Multi-Factor Adaptive E-Learning System. *IAENG Int. J. Comput. Sci.* **2024**, *51*, 1250–1271.
8. Angeioplastis, A.; Aliprantis, J.; Konstantakis, M.; Tsimpiris, A. Predicting Student Performance and Enhancing Learning Outcomes: A Data-Driven Approach Using Educational Data Mining Techniques. *Computers* **2025**, *14*, 83. [CrossRef]
9. Nazempour, R.; Darabi, H. Personalized Learning in Virtual Learning Environments Using Students' Behavior Analysis. *Educ. Sci.* **2023**, *13*, 457. [CrossRef]
10. Murtaza, M.; Ahmed, Y.; Shamsi, J.A.; Sherwani, F.; Usman, M. AI-Based Personalized E-Learning Systems: Issues, Challenges, and Solutions. *IEEE Access* **2022**, *10*, 81323–81347. [CrossRef]
11. Valencia Usme, Y.P.; Normann, M.; Sapsai, I.; Abke, J.; Madsen, A.; Weidl, G. Learning Style Classification by Using Bayesian Networks Based on the Index of Learning Style. In Proceedings of the ECSEE 2023, Seeon, Germany, 19–21 June 2023.
12. Fayaza, M.F.; Ahangama, S. Use of Felder Silverman Learning Style Model in Technology Enhanced Learning. In Proceedings of the 2024 International Research Conference on Smart Computing and Systems Engineering (SCSE), Colombo, Sri Lanka, 4 April 2024; IEEE: Piscataway, NJ, USA, 2024; Volume 7, pp. 1–6.
13. Saputra, J.P.B.; Prabowo, H.; Gaol, F.L.; Hertono, G.F. Identifying student learning styles using support vector machine in the Felder-Silverman learning style model. In Proceedings of the 2023 15th International Congress on Advanced Applied Informatics Winter (IIAI-AAI-Winter), Bali, Indonesia, 11–13 December 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 149–154.
14. Sihombing, J.H.; Laksitowening, K.A.; Darwiyanto, E. Personalized e-learning content based on felder-silverman learning style model. In Proceedings of the 2020 8th International Conference on Information and Communication Technology (ICoICT), Yogyakarta, Indonesia, 24–26 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.

15. Sensuse, D.I.; Hasani, L.M.; Bagustari, B. Personalization strategies based on Felder-Silverman learning styles and its impact on learning: A literature review. In Proceedings of the 2020 3rd International Conference on Computer and Informatics Engineering (IC2IE), Yogyakarta, Indonesia, 15–16 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 293–298.
16. Buciuman, C.F.; Potra, S. Revolutionizing Education in Industry 4.0: Eye-Tracking and AI for Personalized Learning. *Procedia Comput. Sci.* **2025**, *253*, 1658–1667. [CrossRef]
17. Nafea, S.M.; Siewe, F.; He, Y. On Recommendation of Learning Objects Using Felder–Silverman Learning Style Model. *IEEE Access* **2020**, *8*, 125092–125107. [CrossRef]
18. Fernando-Raguro, M.C.; Carpio Lagman, A.; Abad, L.P.; Ong, P.L.S. Extraction of LMS Student Engagement and Behavioral Patterns in Online Education Using Decision Tree and K-Means Algorithm. In Proceedings of the APIT 2022, Virtually, 24–26 February 2022; ACM: New York, NY, USA, 2022; pp. 138–143.
19. Hasibuan, M.S.; Aziz, R.Z.A.; Dewi, D.A.; Kurniawan, T.B.; Syafira, N.A.S. Recommendation Model for Learning Material Using the Felder–Silverman Learning Style Approach. *HighTech Innov. J.* **2023**, *4*, 811–822. [CrossRef]
20. Prabpal, S.; Nitiwatthana, K. Enhancing Learning Style Identification through Advanced Machine Learning Techniques. *Sociolytics J.* **2024**, *1*, 21–33.
21. El-Bishouty, M.M.; Aldraiweesh, A.; Alturki, U.; Tortorella, R.; Yang, J.; Chang, T.W.; Graf, S.; Kinshuk. Use of Felder and Silverman learning style model for online course design. *Educ. Technol. Res. Dev.* **2019**, *67*, 161–177. [CrossRef]
22. Zlatkovic, D.; Denic, N.; Petrovic, M.; Ilic, M.; Khorami, M.; Safa, A.; Wakil, K.; Petković, D.; Vujičić, S. Analysis of Adaptive E-Learning Systems with Adjustment of Felder–Silverman Model in a Moodle DLS. *Comput. Appl. Eng. Educ.* **2020**, *28*, 803–813. [CrossRef]
23. Zagulova, D.; Boltunova, V.; Katalnikova, S.; Prokofyeva, N.; Synytsya, K. Personalized E-Learning: Relation Between Felder–Silverman Model and Academic Performance. *Appl. Comput. Syst.* **2019**, *24*, 25–31. [CrossRef]
24. Kaur, R.; Gupta, D.; Madhukar, M.; Singh, A.; Abdelhaq, M.; Alsaqour, R.; Breñosa, J.; Goyal, N. E-Learning Environment Based Intelligent Profiling System for Enhancing User Adaptation. *Electronics* **2022**, *11*, 3354. [CrossRef]
25. Batool, S.; Rashid, J.; Nisar, M.W.; Kim, J.; Kwon, H.Y.; Hussain, A. Educational data mining to predict students' academic performance: A survey study. *Educ. Inf. Technol.* **2023**, *28*, 905–971. [CrossRef]
26. Nahar, K.; Shova, B.I.; Ria, T.; Rashid, H.B.; Islam, A.S. Mining educational data to predict students performance: A comparative study of data mining techniques. *Educ. Inf. Technol.* **2021**, *26*, 6051–6067. [CrossRef]
27. Ratner, B. *Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2017.
28. Gonzales, L.K.; Glaser, D.; Howland, L.; Clark, M.J.; Hutchins, S.; Macauley, K.; Close, J.F.; Leveque, N.L.; Failla, K.R.; Brooks, R.; et al. Assessing learning styles of graduate entry nursing students as a classroom research activity: A quantitative research study. *Nurse Educ. Today* **2017**, *48*, 55–61. [CrossRef]
29. Green, A.J.; Sammons, G.E. Student learning styles: Assessing active learning in the hospitality learners model. *J. Hosp. Tour. Educ.* **2014**, *26*, 29–38. [CrossRef]
30. Karim, F.; Majumdar, S.; Darabi, H.; Harford, S. Multivariate LSTM-FCNs for time series classification. *Neural Netw.* **2019**, *116*, 237–245. [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.

computers

MDPI

Article

Virtual Reality for Hydrodynamics: Evaluating an Original Physics-Based Submarine Simulator Through User Engagement

Andrei-Bogdan Stănescu [1,*], Sébastien Travadel [2] and Răzvan-Victor Rughiniș [1]

[1] Faculty of Automatic Control and Computers, National University of Science and Technology "Politehnica" Bucharest, 060042 Bucharest, Romania; razvan.rughinis@cs.pub.ro
[2] Collège des Sciences Navales, MINES Paris—PSL, 75272 Paris, France; sebastien.travadel@minesparis.psl.eu
* Correspondence: andrei.stanescu23@gmail.com

Abstract: STEM education is constantly seeking innovative methods to enhance student learning. Virtual Reality technology can represent a critical tool for effectively teaching complex engineering subjects. This study evaluates an original Virtual Reality software application, entitled *Submarine Simulator*, which is developed specifically to support competencies in hydrodynamics within an Underwater Engineering course at MINES Paris—PSL. Our application uniquely integrates a customized physics engine explicitly designed for realistic underwater simulation, significantly improving user comprehension through accurate real-time representation of hydrodynamic forces. The study involved a homogeneous group of 26 fourth-year engineering students, all specializing in engineering and sharing similar academic backgrounds in robotics, electronics, programming, and computer vision. This uniform cohort, primarily aged 22–28, enrolled in the same 3-month course, was intentionally chosen to minimize variations in skills, prior knowledge, and learning pace. Through a combination of quantitative assessments and Confirmatory Factor Analysis, we find that Virtual Reality affordances significantly predict user flow state (path coefficient: 0.811) which then predicts user engagement and satisfaction (path coefficient: 0.765). These findings show the substantial educational potential of tailored Virtual Reality experiences in STEM, particularly in engineering, and highlight directions for further methodological refinement.

Keywords: STEM education; VR; learning program; cognitive process

1. Introduction

Virtual Reality (VR) software applications for learning Science, Technology, Engineering, and Mathematics (STEM) subjects offer a transformative approach to education by providing immersive, interactive, and engaging environments that enhance the process of understanding complex and abstract concepts. VR's capability to simulate real-world scenarios and environments that are otherwise inaccessible makes it a valuable tool in STEM education.

1.1. Existing Tools and Frameworks

Various VR platforms and frameworks have been developed for STEM education, including ScienceVR for science laboratories [1], Unity-based modules for K-12 education [2], and smartphone-based VR plotting systems [3]. These tools enable the creation of 3D visualizations and interactive simulations, enhancing students' understanding of complex concepts. Studies have demonstrated that VR can improve learning efficiency, especially when incorporating environmental traversal capabilities [4].

VR has been effectively integrated across diverse STEM disciplines in higher education, including engineering and chemistry experiments, where it has demonstrated significant gains in spatial abilities, knowledge retention, student motivation, engagement, and overall learning outcomes [5,6].

Mobile VR platforms enhance the affordability and accessibility of VR technology for educational institutions, facilitating its widespread adoption in instructional practices [5,7]. To further support this, VR software must remain user-friendly and compatible with budget-conscious hardware solutions, such as head-mounted displays and haptic gloves, which institutions prioritize for cost-effectiveness [8].

Virtual laboratories have garnered significant attention in research, offering a secure alternative for experiments that might endanger participants. By generating immersive three-dimensional digital environments, VR enables students to engage with and investigate intricate STEM concepts in a risk-free, controlled manner—proving especially advantageous for hazardous or impractical real-world procedures [6,9]. Moreover, the heightened sense of presence and interactive learning via physical movements and gestures in a VR setting can enrich the educational process, rendering abstract ideas more concrete and comprehensible [10].

1.2. Research Gap

Difficult challenges remain in designing effective VR experiences for education, with researchers proposing design principles and guidelines to optimize learning outcomes [10–13]. Current literature lacks comprehensive studies on the affordances of VR for fostering flow states in hydrodynamics learning, leaving educators without evidence-based guidelines to integrate such immersive simulations into curricula.

While numerous VR tools support general STEM learning—such as those for molecular manipulation, ecosystem exploration, or basic physics experiments—few integrate realistic water simulations with intuitive interfaces that enable rapid skill acquisition, particularly for complex underwater scenarios. For instance, projects like Stanford's VR simulation for ocean acidification provide educational dives into marine environments, but they often prioritize visualization over dynamic, physics-accurate interactions [14].

Similarly, underwater VR modules for marine science courses, developed through US National Science Foundation (NSF) funded initiatives, focus on immersive labs but may lack the fidelity needed for engineering-specific tasks like manipulating submerged objects under realistic fluid forces [15,16].

Existing educational VR tools frequently rely on generic physics engines (e.g., those in Unity or Unreal Engine), which are optimized for real-time performance in games but fall short in accurately modeling complex underwater forces such as buoyancy, drag, turbulence, and viscous flow [17,18].

A comparative study of physics engines like ODE, Bullet, and MuJoCo reveals limitations in handling industrial or underwater robotics simulations, where inaccuracies in collision detection, force application, and environmental realism can lead to deviations from expected behaviors [19].

These engines often prioritize computational efficiency over precision, resulting in simplified approximations that do not capture the full complexity of underwater dynamics, such as variable water densities or multi-body interactions in submerged environments. For example, in underwater robotic simulators, generic engines struggle with sensor fidelity and sim-to-real transfer, leading to runtime failures or unrealistic outcomes that undermine training effectiveness [19].

This gap significantly limits the educational value of VR in advanced engineering courses, where high-fidelity simulations are essential for developing intuition in non-intuitive settings like underwater operations [20].

Research on high-fidelity modeling for submarines and remotely operated vehicles (ROVs) demonstrates that custom, physics-based simulations, which often use specialized tools like GPU-accelerated renderers or dedicated emulators, better replicate real-world conditions, improving skill transfer and decision-making in virtual training [21,22].

Expanding on this, bridging the gap could involve hybrid approaches such as integrating deep reinforcement learning for adaptive water flow simulations with user-friendly interfaces to foster rapid mastery [23]. Such advancements would not only boost immersion and flow states in learning but also prepare students for real-world challenges in fields like marine engineering, where a precise understanding of hydrodynamic forces is critical for innovation and safety [24,25].

Overall, literature underscores the need for more specialized VR tools to elevate STEM education from generic experiences to targeted, high-impact simulations that mirror professional demands. Addressing these shortcomings, our application introduces a specialized, custom-developed physics engine and software application—*Submarine Simulator*—explicitly tailored for hydrodynamics and submarine modeling, significantly enhancing realism and practical applicability by incorporating advanced fluid dynamics principles and precise force modeling that are being overlooked by generic engines.

This custom approach allows for more accurate replication of submarine behaviors, such as multi-body interactions in variable water densities, which supports better sim-to-real transfer and decision-making in virtual training environments.

Additionally, the software is engineered for rapid skill acquisition and an efficient, iterative design process. By emphasizing intuitive, high-fidelity interactions, our solution not only addresses the fidelity gap in underwater engineering education but also fosters active learning, significantly enhancing retention and practical applications of complex engineering concepts in real-world underwater environments.

To conclude, despite the rapid advancements in VR technology and education, there remains a significant gap in empirical research examining how VR applications enhance student engagement and understanding in specialized fields like underwater engineering, particularly in simulating hydrodynamics and 3D modeling in VR.

The urgency of addressing this gap is underscored by the escalating demand for innovative STEM education tools amid global challenges such as climate change and ocean exploration, where effective training in underwater engineering can accelerate real-world innovations and sustainability efforts. Underwater engineering plays a critical role in designing sustainable offshore structures, renewable energy systems (e.g., tidal and wave energy), and submersible technologies for environmental monitoring.

1.3. Aim of the Research

In our research, we aim to assess our original VR software designed for acquiring knowledge in hydrodynamics through self-reported engagement from the users. The information provided by the software enables the assimilation of the necessary knowledge for understanding the basic principles of building a small-scale submarine and experimenting with how the shape and properties of the ensemble affect the way it behaves in a simulated underwater environment.

To reach the aim of the research, the software has been tested during a course on underwater engineering at MINES Paris—PSL, offering students an additional opportunity to prototype and iterate on their ideas digitally. Gaining a deep understanding of the unique characteristics of the VR-based learning process is essential, as it empowers students to

actively contribute to the iterative refinement of educational software aimed at fostering specialized expertise.

It is important to note that this study does not seek to benchmark the learning efficacy of VR-based techniques against conventional instructional methods. Instead, it assesses the ability of the original VR system designed for underwater engineering education to evoke targeted psychological and cognitive responses.

The main contributions of this paper are the empirical exploration of VR in enhancing underwater engineering education within challenging, non-intuitive environments that are difficult to visualize mentally. Through a custom-developed VR application, the study addresses three key research questions:

RQ1: How do the affordances of custom-developed VR applications in engineering education impact users' state of flow?

RQ2: In what ways does a user's perceived state of flow influence the overall perceived value of VR-based learning environments?

RQ3: What is the relationship between VR affordances, user flow states, and the effectiveness of engineering education as evaluated through perceived learning value?

2. Features of VR for STEM Education

To ensure an effective learning environment, the VR software should accomplish a set of criteria, aiming to ensure the standard for an educational tool. The current research highlights key features that VR software should include. It also suggests approaches for designing new tools in this field. These features are crucial for motivating students and improving their learning outcomes in STEM disciplines [5]. Efficient VR software for STEM education is characterized by its ability to create high levels of immersion, interactivity, and engaging learning environments [10].

2.1. Immersive Learning Environments

Immersive features of VR software are pivotal in creating engaging and realistic experiences across various domains, including STEM. Immersive VR can create a strong sense of presence and boost engagement. Engagement in educational contexts can be defined as the sustained cognitive, emotional, and behavioral involvement of users, achieved through immersion that captivates senses, interactivity that empowers actions, and engaging learning environments that provide meaningful context and motivation. The immersive nature of VR is achieved through a combination of technological and design elements that simulate real-world experiences in a virtual environment. These features leverage advanced visualization, interaction, and sensory integration to enhance user engagement and understanding. Additionally, the design of VR environments must ensure sufficient fluidity and immersion to avoid negatively impacting learning outcomes [5,26].

VR's immersive features allow for personalized learning experiences, helping students understand and explore complex and abstract concepts that are often difficult to replicate in traditional educational settings through interactive and engaging methods [27]. VR is used in training scenarios to replicate real-world tasks, utilizing scripting and game engines to create customizable and adaptive training environments [28].

The effectiveness of immersive features in educational VR applications is mixed, highlighting the need for careful alignment of design features with educational goals [29]. While immersive features in VR software offer significant potential for enhancing user experiences, they also present challenges that need to be addressed. VR applications should provide a high degree of immersion, allowing students to feel present within the virtual environment. This is essential for engaging students and enhancing their learning

experience [10,26]. The effectiveness of VR in translating immersive experiences into academic performance requires further exploration [30].

2.2. Interaction: 360-Degree Video and Head-Mounted Displays

Notably, 360-degree environments and head-mounted display technologies are used for VR experiences, allowing viewers to explore environments from all angles. As a result, it can increase emotional engagement and empathy [31].

VR software leverages high-resolution graphics and 3D modeling to craft realistic environments with genuine depth. This contrasts with 360-degree videos, which lack depth and thus offer only a limited immersion experience. Nonetheless, the benefits of true 3D VR are apparent in applications like digital media and interactive art, where deep learning models elevate the immersive experience by boosting modeling accuracy and the sense of presence [32]. Platforms like CAVE and head-mounted displays provide a profound sense of immersion by creating a neuropsychological sense of "*being there*" [33].

Multi-sensory virtual environments, such as the Multi-Sensory Virtual Decision-Making Center, facilitate collaborative learning by allowing multiple users to engage in real-time decision-making [34]. This approach not only improves the educational experience but also enhances teamwork and communication skills among students.

The design of VR experiences must be carefully aligned with the intended outcomes to maximize their effectiveness. As VR technology continues to evolve, it is expected that these challenges will be mitigated, leading to widespread adoption and innovation across various fields.

2.3. Interactive Learning

VR applications, such as educational games and simulations, have been shown to increase student motivation and engagement by making learning more interactive and enjoyable [6,35]. Applications should enable students to interact with virtual objects and environments, facilitating hands-on learning and exploration. This can be achieved through advanced interfaces like hand gesture recognition [5,36]. VR applications should offer customizable environments that cater to individual learning needs, allowing students to explore topics at their own pace and according to their interests [26].

The promotion of active learning through movement and gestures in a three-dimensional virtual environment can lead to a deeper understanding and retention of knowledge [5]. Effective Human-Computer Interaction (HCI) is central to immersive experiences, allowing users to interact with virtual environments seamlessly [37]. In educational settings, VR applications often include interactive elements that allow users to manipulate and explore abstract concepts, although narrative and social features are less commonly integrated [29].

The software should adapt to the learner's progress, providing tailored feedback and challenges to optimize learning outcomes [38,39]. It should be scalable to accommodate different educational settings, from individual learners to large classrooms [40], and it should support collaborative learning, enabling students to work together in virtual spaces, share ideas, and solve problems collectively [41]. Incorporating social elements can enhance engagement and motivation, making learning more enjoyable and effective [39]. While VR offers significant advantages for STEM education, challenges such as the need for teacher training and the integration of VR into existing curricula must be addressed.

2.4. Gamification in STEM VR Education

Different game-design elements within VR environments have the role of enhancing learning experiences in science, technology, engineering, and mathematics. This approach aims to increase student engagement, motivation, and knowledge retention. The use of

gamification and personalized learning in VR environments can further captivate students' attention and encourage exploration and discovery [42].

Common gamification mechanisms in VR include rewards, challenges, and avatars. Elements such as content unlocking, point systems, task difficulty levels, and achievement systems are frequently used to create engaging learning environments. These elements are designed to motivate students by providing immediate feedback and a sense of progression [43].

VR platforms like VRCoding provide interactive and immersive experiences that encourage critical thinking and problem-solving. These environments allow students to actively participate in their learning process, moving away from passive listening to engaging with the content in a meaningful way [44].

Studies have shown that gamified VR environments significantly improve student engagement and motivation. For instance, the use of a gamified 3D virtual world for teaching computer architecture resulted in higher engagement levels and improved learning outcomes among engineering students [45]. While gamification can enhance learning, it is essential to balance game elements to avoid negative impacts on intrinsic motivation. For example, the use of high scores and achievements can lead to increased competition, which may detract from the learning experience [46].

3. Original VR Software to Teach Hydrodynamics

Drawing upon the previously mentioned elements, a tailored software application was developed to support the learning of core hydrodynamics principles essential for submarine design.

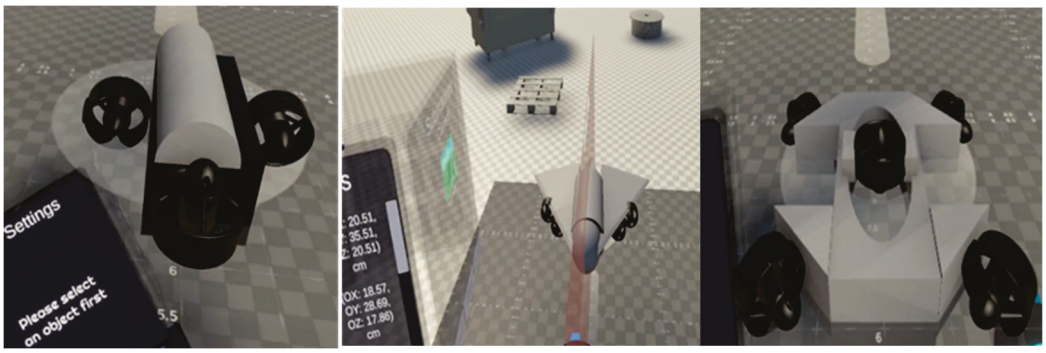
Submarine Simulator is a custom-made application created in Unity that facilitates engineering education both in VR and Augmented Reality (AR). The application consists of three main modes:

- *Environment Selection*: Once immersed in VR, students can choose a virtual environment before beginning the construction process of a small-scale submarine model. The application currently allows students to choose from immersing themselves in an open world, a garage, a hangar, or an open space in nature. At this stage, students have the option to switch to AR instead of VR. This allows them to perceive their surrounding physical environment while utilizing a real table as a blended digital-physical canvas to initiate the construction process (Figure 1).

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

Figure 1. Using the main menu during construction. (**a**) Selecting 3D shapes. (**b**) Setting the properties of a 3D shape.

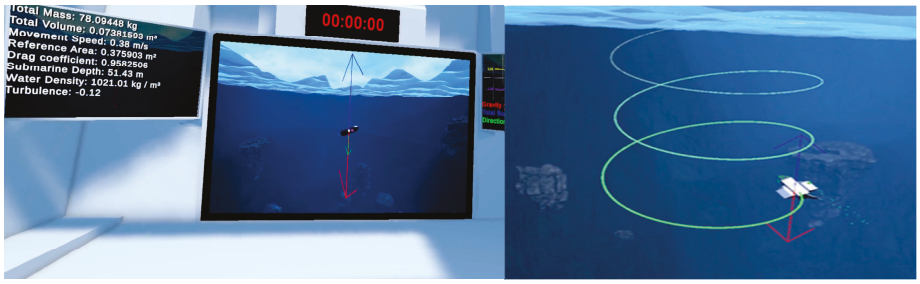
- *Construction Scene*: After picking their favorite virtual environment, the students can use the VR controllers to drag, drop, and combine different 3D shapes to create a submarine model (Figure 2). They have multiple construction instruments and functions that allow for building with precision. Additionally, the students can change the weight, density, and material of each of the shapes added to the environment. Once the students decide that the shape of the submarine is complete, they may also attach motors. If the student has selected to use AR, all activities will be made by manipulating holograms which will blend together with the real space around them.



(**a**) (**b**) Model 2 (**c**) Model 3

Figure 2. Three different submarine models showing the flexibility of the construction process. (**a**) Model 1. (**b**) Model 1. (**c**) Model 1.

- *Simulation Scene*: After completing the submarine model, students can test it in a simulated underwater scene (Figure 3). Based on preference, they can control the submarine either from a third-person or first-person perspective. The scene includes a hyper-realistic water simulator that supports the player in understanding how the previously created model behaves underwater. Using the motors attached previously, they can control the submarine, navigate underwater, and visually understand how the thrust and the physical characteristics of the submarine model interact with the water environment.

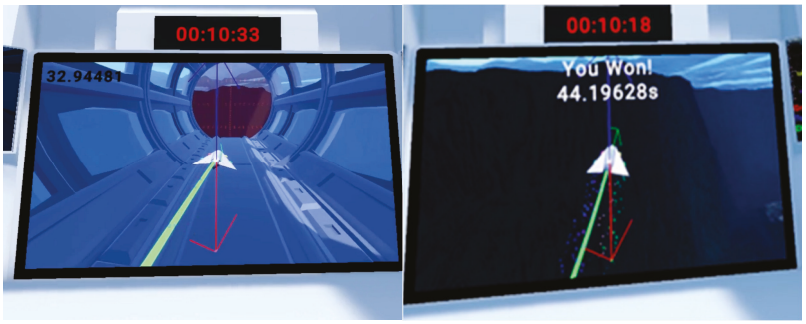


(**a**) (**b**)

Figure 3. The Underwater Simulation Mode. (**a**) Command dashboard in simulation mode. (**b**) Spiraling model in simulation mode.

- *Gamification*—Gamification in VR has been linked to better knowledge retention and understanding (Figure 4). The integration of game elements helps students grasp complex concepts more effectively, as seen in the positive feedback from students using the VR Coding system for learning computational thinking [44–46]. Our setup leverages gamification principles—such as points, leaderboards, real-time competition, and immediate feedback—to transform the learning process into an interactive and motivating experience. These principles deepen comprehension by encouraging

problem-solving in a dynamic VR context, where failures and successes provide direct learning opportunities.



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

Figure 4. Use of gamification in the Submarine Simulator. (**a**) Submarine model competing against other students in underwater track. (**b**) Submarine model winning a race.

In our testing process, participants had the opportunity to test their custom-designed 3D submarine models by competing in virtual underwater racetracks, all simulated within the immersive VR environment. Upon successfully completing a race, students earned points based on their performance and could view their final completion time for the specific track. To enhance engagement, the system supported real-time multiplayer racing, where users could compete against their colleagues. This was facilitated through a visual "shadow" representation of peers' submarines, allowing participants to see and react to others' progress dynamically throughout the course. Ultimately, students achieving the highest cumulative scores across races were declared winners, fostering a competitive and collaborative learning atmosphere.

It is worth mentioning that the instructor overseeing these learning sessions could monitor everything live. The virtual simulation viewed via the student's headset was duplicated on the instructor's displays, enabling him to provide appropriate guidance and assistance to the participants if required.

4. Methodology

The assessment of the value inherent in VR software requires a comprehensive methodology that encompasses various dimensions, including performance, usability, user engagement, and the context in which the software is being used. The evaluation process is very important for identifying potential challenges and for ensuring that the software effectively fulfills its intended objectives.

4.1. Research Objectives and Hypothesis

The primary objective of this study was to evaluate the perceived effectiveness of a custom-built software application, termed *Submarine Simulator*, within the context of a course on underwater engineering in the fourth year of the engineering bachelor's program at MINES Paris—PSL, France.

Other objectives were related to assessing the software's impact on student engagement, specifically exploring how its tailored affordances influence the educational experience in a specialized engineering curriculum. By integrating advanced simulations, the research sought to determine the extent to which the *Submarine Simulator* enhances students' understanding of complex underwater engineering concepts and fosters an immersive learning environment conducive to achieving a state of flow.

More specifically, we tested the following hypotheses:

H1. *Affordances provided by custom-developed VR applications for engineering education significantly influence users' reported state of flow.*

H2. *The perceived state of flow among users significantly affects the perceived value of the learning environment, as measured by the CEGE framework.*

4.2. Research Methods

4.2.1. The Quality of the Experience

To assess the engagement and quality of the VR experience, a custom-made affordance questionnaire was administered immediately following the first practice session (Phase 1). Affordances, as defined by Gibson (1979), refer to the possibilities for action provided by an environment, which in the context of VR systems like the *Submarine Simulator*, include how intuitive, accessible, and engaging the interface and interactions are for users [47]. The questionnaire was developed based on general guidelines for affordance evaluation, ensuring it captures the specific affordances relevant to the VR environment.

The affordance questionnaire consists of 20 items across four key dimensions:

- Actions: This item measures the clarity and intuitiveness of interaction possibilities (e.g., ease of navigating the VR environment).
- Engagement: Assesses the degree to which the VR system fosters immersive and sustained user involvement.
- Fluid Mechanics: Evaluates the smoothness and responsiveness of dynamic interactions (e.g., object manipulation, movement within the simulator).
- Functions: This item measures the effectiveness and accessibility of the VR software's core building and maneuvering functionalities (e.g., control panels, functions).

Each item is rated on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree), allowing for nuanced quantification of user perceptions. The questionnaire, detailed in Appendix A, was administered immediately after their Phase 1 session to capture real-time feedback and minimize recall bias. The greater the total score, the stronger the perceived affordance of the VR system.

4.2.2. The Flow State Scale (FSS) Questionnaire

Developed by Jackson and Marsh (1996), the FSS is a validated psychometric tool designed to measure the psychological state of flow, characterized by deep immersion, focus, and enjoyment in an activity [48].

The FSS was administered immediately following the second practice session (Phase 2) to capture participants' flow experiences while interacting with the VR software. The questionnaire consists of 36 items, grouped into nine subscales, with the dimensions being measured on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree):

- Challenge-Skill Balance: Perceiving that personal skills match the task's demands.
- Action-Awareness Merging: Experiencing seamless integration of actions and awareness.
- Clear Goals: Having a clear understanding of objectives.
- Unambiguous Feedback: Receiving immediate, clear feedback on performance.
- Concentration on the Task: Maintaining deep focus without distractions.
- Sense of Control: Feeling in command of the activity.
- Loss of Self-Consciousness: Becoming less aware of self and external judgments.
- Transformation of Time: Perceiving time as altered, either speeding up or slowing down.
- Autotelic Experience: Finding the activity intrinsically rewarding.

For each of the nine dimensions of the FSS, the scores from the four corresponding items are summed and averaged to produce a mean score. This approach quantifies

participants' flow experiences, enabling the examination of their connections to learning styles and performance outcomes in immersive VR environments, such as the *Submarine Simulator*, where sustained engagement is critical for effective learning.

4.2.3. The Core Elements of Gaming Experience (CEGE) Questionnaire

Developed by Calvillo-Gámez et al. (2010), CEGE is a validated tool designed to measure the quality of a gaming experience by assessing key psychological and interactive elements [49].

It was administered at the end of the third practice session (Phase 3) to evaluate the perceived value of the Submarine Simulator's VR environment in fostering an engaging and rewarding gaming experience. The CEGE questionnaire comprises 38 items grouped into six core dimensions:

- Enjoyment: Measures the level of fun and satisfaction derived from the VR experience.
- Control: Assesses the user's sense of control and ability to interact effectively with the VR system.
- Immersion: Evaluates the depth of psychological absorption in the VR environment.
- Challenge: Gauges the perceived difficulty and stimulation provided by the VR tasks.
- Focus: Measures the state of optimal engagement where users are fully focused and immersed.
- Game Mechanics: Assesses the quality and intuitiveness of the VR system's interactive elements (e.g., controls, in-game construction tools).

Each item is rated on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree), allowing for a nuanced quantification of the gaming experience. CEGE was chosen for its comprehensive coverage of gaming experience facets, which align with the study's goal of evaluating how the *Submarine Simulator's* affordances—actionable properties like intuitive controls and immersive tasks—contribute to perceived value and engagement.

The higher the overall CEGE score (sum of all items, ranging from 38 to 190), the stronger the perceived affordances and value of the VR system.

From both a contextual and stakeholder perspective, understanding the values, needs, and objectives of diverse stakeholders (such as educators, learners, and developers) is critical in the design and evaluation of VR applications [50]. This approach ensures that the software aligns with its intended educational goals, delivering an effective and engaging user experience.

For educational VR applications, evaluation methods often include the Technology Acceptance Model and flow tests to assess how well the software supports learning objectives. In medical education contexts, such as surgical training or stroke rehabilitation simulations, these can be complemented by specialized tools that evaluate the tolerability of artificial movement types (e.g., teleportation or floating movements) to minimize simulator sickness, ensuring safer and more effective VR adoption. This approach involves isolating movement factors in a controlled virtual environment and using validated measures like the Simulator Sickness Questionnaire (SSQ) for pre- and post-exposure assessments, as demonstrated in recent software developments aimed at reducing adverse effects for vulnerable users [51]. In our case, we are assessing immersivity through affordance, FSS, and CEGE. Additionally, we use a refined way of navigating through the environment using a zoom-in and zoom-out motion using the controllers in order to mitigate simulator and motion sickness.

4.3. Statistical Methods

A Confirmatory Path Analysis (CFA) was conducted to determine the relationship between the software features, as evaluated by students, and other variables connected to their psychological states, by using the software SmartPLS (v3.2.9) [52,53]. It provides a

coherent explanatory model of causal mechanisms, highlighting both the direct and indirect effects of independent variables on a dependent variable within a theoretically specified network of relationships.

Partial Least Squares Structural Equation Modeling (PLS-SEM), which is particularly well-suited for exploratory research involving small sample sizes, non-normally distributed data, and formative or composite measurement models, was conducted using the specified software, incorporating bootstrapping with 5000 samples to estimate path coefficients. PLS-SEM offers a nonparametric, variance-based alternative that can yield meaningful results with samples as small as 20–30 inputs, provided model complexity is moderate [53,54]. PLS-SEM does not impose distributional assumptions and utilizes bootstrapping procedures to estimate significance, making it an appropriate choice when data normality or scale properties are not guaranteed. As a result, these features made SmartPLS the most appropriate tool for our exploratory analysis under the conditions of limited sample size and emerging constructs in a novel educational context [54].

Beyond R^2, we also reported f^2 effect sizes, which quantify the contribution of each independent variable to the dependent variable, thereby complementing R^2 with a measure of local effect strength [55]. Additionally, the quality of the model was assessed using model fit indices, specifically SRMR (Standardized Root Mean Square Residual), NFI (Normed Fit Index), d_ULS (Unweighted Least Squares Discrepancy), d_G (Geodesic Discrepancy) and Chi-square values, which serve a similar function to loss/error metrics in predictive modeling, particularly in variance-based SEM contexts. Discriminant validity was evaluated using the Fornell–Larcker Criterion, while model fit was assessed through standardized metrics such as SRMR, NFI, Chi-square, d_ULS, and d_G indices.

While traditional loss functions such as MSE (Mean Squared Error) or cross-entropy are not directly applicable in SmartPLS's estimation framework—given its nonparametric and component-based nature, the bootstrapping procedure (5000 resamples) and outer model loadings offer robust estimates of parameter stability and indicator reliability.

4.4. Target Group

The study involved a selected group of 26 fourth-year engineering students from MINES Paris—PSL, all specializing in engineering and sharing similar academic backgrounds in robotics, electronics, programming, and computer vision. This homogenous cohort, primarily aged 22–28 and enrolled in the same university program, was intentionally chosen to minimize variations in skills, prior knowledge, and learning pace. This homogeneity enabled a more consistent and reliable evaluation of the VR *Submarine Simulator*'s effectiveness, as differences in outcomes could be attributed to the software itself rather than varied participant characteristics.

All 26 students voluntarily participated in the evaluation of the VR Submarine Simulator. To ensure participant anonymity, no personal identifiers, such as names, were used at any point during the study. Instead, each student was assigned a unique numerical code for reference. Every participant successfully completed all three phases of the study, as outlined in the research design (see Section 4.5), ensuring a comprehensive assessment of the VR software.

4.5. Research Design

The study was structured as a series of three sessions, logically divided into three phases, each designed to progressively build students' proficiency in using the VR Submarine Simulator while deepening their understanding of underwater engineering principles.

Conducted over five weeks, these sessions involved distinct tasks tailored to enhance both technical skills and conceptual knowledge. Each of the 26 students completed three

distinct phases. Phases 2 and 3 were conducted approximately one week after the preceding phase, ensuring a consistent schedule for all participants.

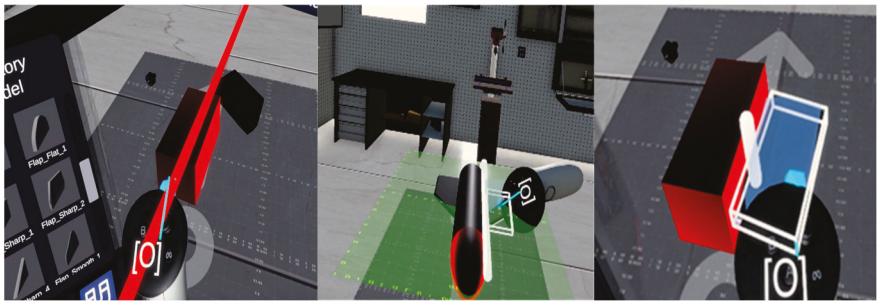
This study was embedded within a 3-month, on-site course dedicated to the design and prototyping of mid-sized underwater remotely operated vehicles (ROVs) ranging from 50 to 150 cm in length.

The figures in this section are direct screenshots from the original *Submarine Simulator* VR software. The following phases were designed and executed by each of the participants in our study:

4.5.1. Phase 1: Introduction and Foundational Training

This phase has been split into four chapters:

- Chapter 0—Video Introduction: All students have been introduced to the capabilities of the VR software using a 10-min video explanation showing all functionalities available and numerous examples.
- Chapter 1—Building Mode: All students were invited to engage in the immersive VR experience, where they completed a step-by-step tutorial led by a trained facilitator. This hands-on session introduced them to the essential functions needed to construct small-scale submarine models, fostering a clear understanding of the VR *Submarine Simulator's* core features (Figure 5).
- Floatability Testing and Underwater Navigation: After mastering the construction process, students were invited to a virtual underwater environment where they could test their submarine models. This interactive setting allowed them to observe and analyze key hydrodynamic principles, including floatability, buoyancy, acceleration, and other dynamic behaviors, as applied to their designs.
- Exploratory Learning: Lastly, once showing a basic level of understanding and comfort while using the VR software, all students were given a construction task. The students had complete creative freedom during the construction process.



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

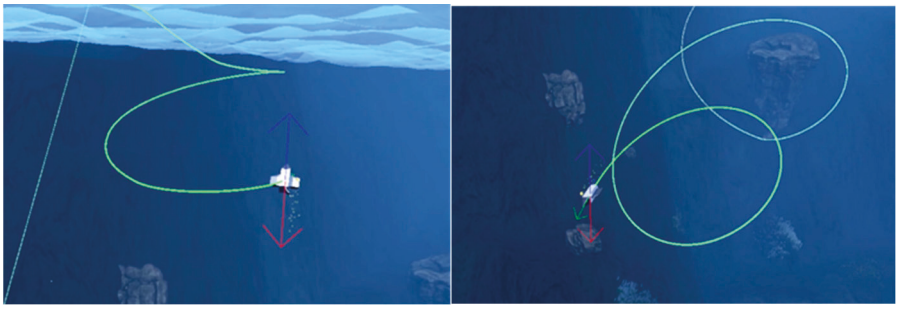
Figure 5. Submarine designing process in VR environment using different functionalities. (**a**) Rescaling objects. (**b**) Combining objects. (**c**) Centering objects.

The VR session was limited to 35 min to prevent strain from participants' unfamiliarity with the technology. A comprehensive list of commands and guidelines provided to all students is detailed in Appendix B.

4.5.2. Phase 2: Underwater Design for Spiral Trajectories

Following the re-accommodation in VR, participants were given two construction tasks designed to deepen their understanding of hydrodynamic principles through practical application.

Using the VR simulator's building mode, students were tasked with creating two distinct submarine models: one to navigate in a tight spiral trajectory and another to follow a loose spiral trajectory (Figure 6). The following command has been given explicitly:



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

Figure 6. Examples of models designed by the students of the target group. (**a**) Model in a tight spiral motion. (**b**) Model in a loose spiral motion.

> *"For the next 35 min, your task is to design two unique submarine models using the VR Submarine Simulator. Each model must include at least four shapes and two motors. One submarine should follow a tight spiral path with a radius of approximately 2 m, while the other should move in a loose spiral path with a radius of more than 5 m. You can build them in any order and adjust your designs as many times as needed. Once I confirm that a model's trajectory is correct, you can move on to the next part of the exercise. Have fun creating and experimenting! Any questions before we start?"*

These tasks required students to adjust key parameters, such as buoyancy, weight distribution, and motor positioning, to achieve the desired movement patterns in the simulated underwater environment. The tight spiral task demanded precise control to maintain stability within a confined radius, testing students' ability to balance hydrodynamic forces under constrained conditions. In contrast, the loose spiral task allowed for broader movement, emphasizing sustained control over longer distances and varying currents.

Real-time feedback from the simulator's tailored physics engine, featuring green visual markers for model stability and trajectory accuracy, allowed students to iteratively refine their submarine designs with precision and confidence.

4.5.3. Phase 3: Paired Competition in Underwater Racetracks

In the concluding phase of the study, conducted 7 days after Phase 2, a dynamic competitive element was introduced by organizing the 26 participants into 13 pairs, fostering both collaborative teamwork within each duo and spirited rivalry across teams.

Each pair worked together to strategize and generate creative solutions, then individually crafted submarine models tailored for three progressively demanding virtual underwater race tracks:

- **Track 1:** Focused on straight-line propulsion with stable, deviation-free submarine models (Figure 7).
- **Track 2:** Required precise left/right maneuvers (Figure 8).
- **Track 3:** Demanded full maneuverability (up/down/left/right) in a complex underwater environment (Figure 9).

These meticulously designed tracks simulated real-world engineering scenarios, rigorously evaluating participants' navigation precision, model performance, and design ingenuity in an immersive, gamified environment that heightened engagement and practical learning.

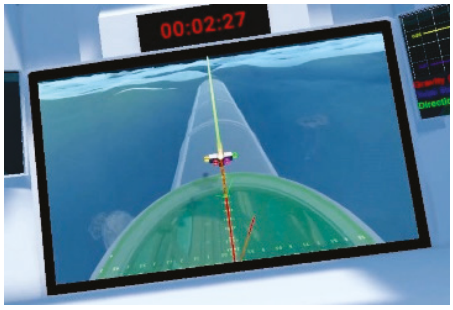


Figure 7. The linear navigation course.

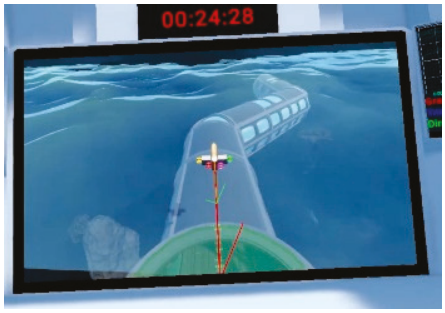


Figure 8. The lateral navigation course.

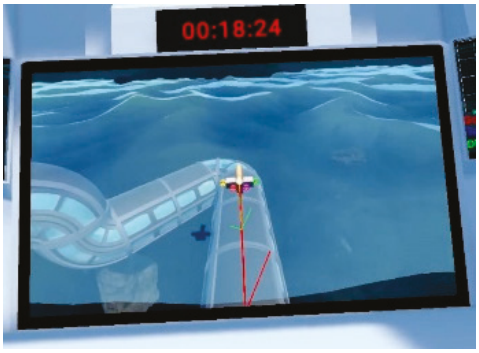


Figure 9. Race track testing model's vertical and horizontal maneuverability.

4.6. Control Conditions

To ensure the reliability and validity of the research, several control measures were implemented to minimize confounding variables and standardize the experimental conditions across all 26 participants. These measures were implemented to isolate the VR software's impact on user engagement, making sure that any observed differences could be attributed solely to the intervention, rather than external variables.

The VR sessions took place in a controlled laboratory at MINES Paris—PSL, with consistent lighting (ambient, 500 lux) and temperature (22–24 °C) to maintain a uniform environment and minimize distractions. No external interruptions were permitted during the sessions, ensuring uninterrupted focus on the tasks.

All participants used standardized hardware configurations, featuring Meta Quest 3 (Meta Platforms, Inc., Menlo Park, CA, USA) VR headsets, paired controllers, and high-performance rendering laptops (HP OMEN 16-inch laptop (HP Inc., Palo Alto, CA, USA) with processor AMD Ryzen 9 6900HX (Advanced Micro Devices, Inc., Santa Clara, CA, USA), and graphics card NVIDIA GeForce RTX 3070Ti (NVIDIA Corporation, Santa Clara, CA, USA)), connected via USB-C cables to ensure minimal latency, eliminate lag, and

prevent variability from equipment differences. Audio levels were consistently set on the headsets across all sessions to maintain a uniform sensory experience.

To standardize participant instructions, a trained facilitator delivered all commands verbally using a scripted protocol (detailed in Appendix B), ensuring consistency in guidance across all sessions. This approach prevented variations in instruction delivery that could influence participant's performance or understanding of the VR software's functionality.

To maintain consistency, participants adhered to a fixed schedule: those assigned a morning slot (e.g., 9:00–10:00 AM) for one phase continued with approximately the same time for subsequent phases. This controlled timing minimized the impact of fatigue or circadian rhythm variations on performance.

Lastly, participants were screened for VR-related motion sickness susceptibility prior to the study, with none reporting significant discomfort, ensuring consistent engagement with the VR environment.

4.7. Definition of Variables

To investigate the connection between user perceptions and the software features of the technical solution, we collected the following variables and their sub-components as critical and relevant for our analysis, based on data collected via questionnaires. These variables, detailed alongside their sub-components and corresponding loading factor values, are presented in the table below (Table 1).

Table 1. Variable description.

Variables	Definition	Sub-Components	Sub-Component Definition	Loading Factors Value
FLOW	Characteristics for the Flow State	Action Awareness	Merging of action and awareness.	0.653
		Autotelic Experience	The activity is its own reward.	0.813
		Challenge Skills	Balance between challenge and skill level.	0.558
		Clear Goals	Knowing what needs to be conducted in the given circumstance	0.466
		Concentration Tasks	Focused attention on the task.	0.645
		Loss of Self-Control	Feeling of effortless action.	0.664
		Paradox Control	Sense of control without deliberate effort.	0.803
		Unambiguity	Immediate feedback on performance.	0.373
AFFORDANCE	Traits Related to Affordance	Engagement	The degree to which a user interacts with or is drawn into the perceived possibilities of an object.	0.483
		FluidMech	How close to reality the fluids and forces behave in the simulation.	0.648
		Actions	The visible or perceptible cues an object provides about its possible actions or uses.	0.595
		Functions	The available functions to manipulate shapes and models.	0.600

Table 1. *Cont.*

Variables	Definition	Sub-Components	Sub-Component Definition	Loading Factors Value
CEGE	Software Evaluation	Control	A sense of agency or influence over the situation/task.	0.333
		Enjoyment	Experiencing pleasure or satisfaction from the activity.	0.971
		Environment	The setting or context in which the activity occurs.	0.395
		Facilitator	Something in the simulation that helps make an action or process easier.	0.723
		Gamification	How the game-design elements have been applied in the simulation.	0.603
		Ownership	Feeling of personal connection or responsibility towards the task/activity.	0.705

5. Results

The data structure and distribution alongside the raw data have been expanded in Appendix C. Each variable was standardized to z-scores, centering and scaling values relative to their distribution.

To answer the research questions, a CFA path factor analysis has been applied, which resulted in a model comprising the formative variables: flow state (FLOW), affordance/usability (AFF), and the reflective variable: the Core Elements of the Gaming Experience (CEGE), with the items presented in Table 1.

The statistical analysis provides robust evidence that the immersive and usability affordances of the VR software substantially influence user engagement and the quality of their learning experience. Specifically, the high path coefficients (0.811 and 0.765) indicate strong predictive relationships, demonstrating practically that intuitive interaction and realism directly contribute to achieving an optimal learning state (flow) among students (Figure 10).

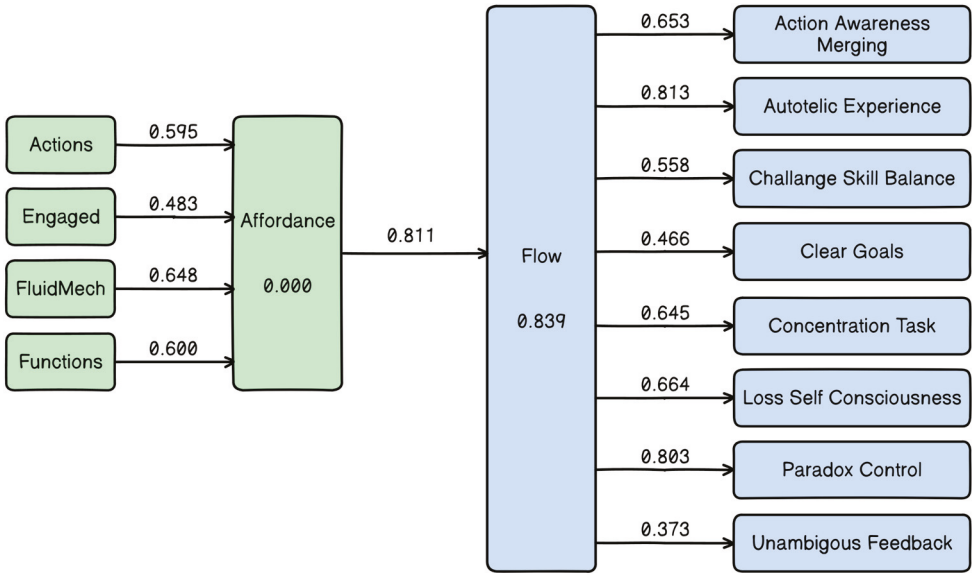


Figure 10. The path coefficients for affordance and flow (Smart PLS output version 3.2.9).

However, some statistical validity measures discussed below (such as AVE below recommended thresholds) suggest that particular survey questions might not fully capture all nuances of user experiences or that responses reflected varied interpretations of questionnaire items. Practically, these results imply that minor refinements in measurement items or additional clarifications during data collection could further strengthen empirical validation of the software's educational benefits.

Overall, we identified that affordances significantly shape the user's state of flow. H1 is confirmed. The very high value of the path coefficient of the model (0.811) proved the relationship.

Table 2 presents the Coefficient of Determination (R^2) that indicates the predictive power of the model, showing the proportion of variance explained by independent variables. Flow (R^2 = 0.658) shows that 65.8% of the variance in flow is explained by affordance, indicating a strong model fit for this relationship. The R^2 Adjusted (0.644) corrects for model complexity and remains high, showing model stability.

Table 2. R^2—Coefficient of Determination for affordance and flow.

Dependent Variable	R^2	R^2 Adjusted	Interpretation (Effect Size)
Flow	0.658	0.644	Strong (Substantial effect, $R^2 > 0.50$)

Table 3 presents the effect size for predictors (f^2), measuring the impact of an independent variable on a dependent variable. Affordance has a very large effect on flow, highlighting that ease of use and immersive design are critical drivers of user engagement. This is an exceptionally strong relationship, indicating that VR affordances significantly shape the user's state of flow.

Table 3. Effect Size for Predictors related to affordance and flow.

Predictor → Dependent Variable	f^2 (Effect Size)	Interpretation (Cohen's Guideline)
Affordance → Flow	1.924	Very Large Effect ($f^2 > 0.35$)

Focus on improving usability, control, and immersion in the VR experience to boost user flow and enjoyment. Table 4 presents reliability and validity metrics for the constructs affordance (AFF) and flow in the VR environment. AFF is a formative variable and no calculus was conducted. Flow is a reflective variable.

Table 4. Construct Reliability and Validity for affordance and flow.

	Cronbach's Alpha (CA)	rho_A	Composite Reliability (CR)	Average Variance Extracted (AVE)
Threshold	≥0.7	≥0.7	≥0.7	≥0.5
AFF	-	1.000	-	-
Flow	0.842	0.862	0.839	0.407

Cronbach's Alpha (CA) for flow: 0.842, indicating good internal consistency. AFF is a formative variable and no calculus was conducted. The Spearman rho_A coefficient for flow: 0.862 shows good reliability. Composite Reliability (CR) for Flow: 0.839 indicates good construct reliability. Average Variance Extracted (AVE) for Flow: 0.407 is below the recommended threshold, meaning it may not explain enough variance in its indicators.

Flow has good internal consistency and reliability (Cronbach's Alpha, rho_A, and CR are all above 0.7). However, Flow's AVE is low (0.407), which suggests potential validity concerns. This could mean that the indicators for flow may not sufficiently capture the construct.

Table 5 presents Discriminant Validity using the Fornell–Larcker Criterion, which checks whether a construct is distinct from others in a model. Square Root of AVE (Diagonal Values): AFF is likely 0.811. Flow: 0.638 (which is the square root of AVE).

Table 5. Fornell–Larcker Criterion.

	AFF	**Flow**
AFF	-	-
Flow	0.811	0.638

The correlation between AFF and Flow (0.811) is strong. Flow's square root of AVE (0.638) is lower than its correlation with AFF (0.811). This suggests poor discriminant validity, meaning flow and affordance might not be clearly distinguishable from one another in the model.

Furthermore, the Fornell–Larcker Criterion suggests insufficient discriminant validity, as the square root of AVE for Flow (0.638) is lower than its correlation with affordance (0.811), which may reflect conceptual overlap between immersive affordances and flow in the interactive VR context. In response, the measurement model should be revised in the future to better delineate these constructs.

In Table 6, the SRMR (Standardized Root Mean Square Residual) value of 0.121 is not very close to the maximum threshold of 0.10, suggesting the model has a low fit with room for improvement.

Table 6. Model Fit for Saturated and Estimated Models on affordance and flow.

	Saturated Model	**Estimated Model**
SRMR	0.121	0.121
d_ULS	1.138	1.138
d_G	0.820	0.820
Chi-Square	81.449	81.449
NFI	0.509	0.509

d_ULS (Unweighted Least Squares Discrepancy) and d_G (Geodesic Discrepancy) lower values indicate a better fit. Chi-Square Value 81.449 for the estimated model is at least equal to the saturated value; thus, we can count on the model fit. NFI = 0.509 < 0.80 indicates a poor fit. Low NFI could be due to sample size issues.

Model fit indices such as SRMR (0.121) and NFI (0.509 and 0.641) fall short of ideal thresholds, suggesting the need for structural refinement and more parsimonious modeling in future analyses.

In the second path analysis (Figure 11), the very high value of the path coefficient of the model (0.765) confirms the supposed relationship between the software characteristics, through flow state, and CEGE. H2 is confirmed as well.

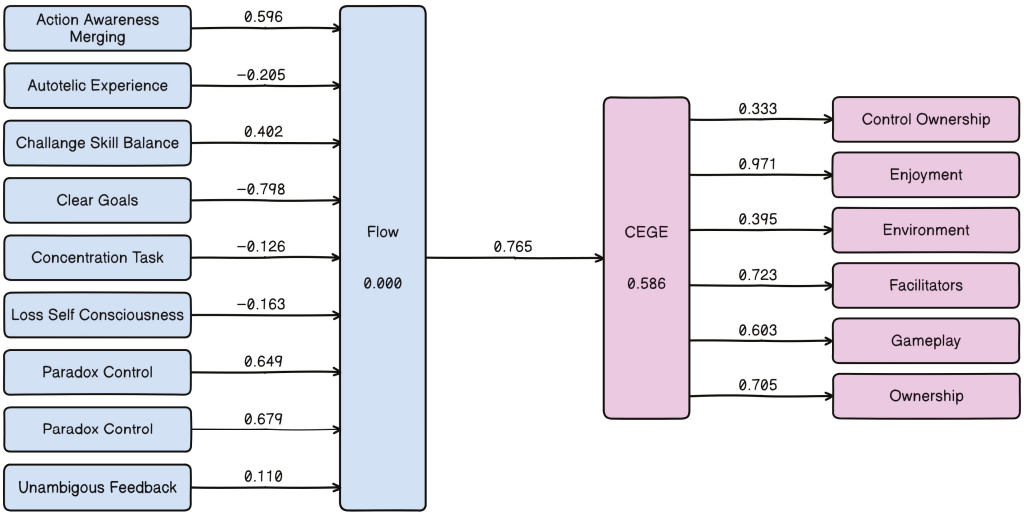


Figure 11. The path coefficients for flow and CEGE (Smart PLS output version 3.2.9).

Table 7 presents the Coefficient of Determination (R^2) that indicates the predictive power of the model, showing the proportion of variance explained by independent variables. CEGE (R^2 = 0.586) shows that 58.6% of the variance in CEGE is explained by flow, indicating a strong model fit for this relationship. The R^2 Adjusted (0.569) corrects for model complexity and remains high, showing model stability.

Table 7. R^2—Coefficient of Determination for flow and CEGE.

Dependent Variable	R^2	R^2 Adjusted	Interpretation (Effect Size)
CEGE	0.586	0.569	Strong (Substantial effect, $R^2 > 0.50$)

Table 8 presents the effect size for predictors (f^2), measuring the impact of an independent variable on a dependent variable. Flow has a very large effect on CEGE, highlighting that ease of use and immersive design are critical drivers of user engagement. This is an exceptionally strong relationship, indicating that VR affordances significantly shape the user's state of flow.

Table 8. Effect Size for Predictors related to flow and CEGE.

Predictor → Dependent Variable	f^2 (Effect Size)	Interpretation (Cohen's Guideline)
Flow → CEGE	1.414	Very Large Effect ($f^2 > 0.35$)

Educators and developers should emphasize enhancements in usability, control, and immersion within the VR experience to elevate user flow and enjoyment. Table 9 presents reliability and validity metrics for the CEGE and Flow (as reflective variables) in the VR environment. AFF is a formative variable and no calculus was conducted.

Cronbach's Alpha (CA) for CEGE: 0.793 indicates good internal consistency. Flow is a formative variable and no calculus was conducted. The Spearman rho_A coefficient for flow, 0.869, shows good reliability. Composite Reliability (CR) for CEGE: 0.803 indicates good construct reliability. Average Variance Extracted (AVE) for CEGE: 0.432 is below the recommended threshold, meaning it may not explain enough variance in its indicators.

CEGE has good internal consistency and reliability (Cronbach's Alpha, rho_A, and CR are all above 0.7). However, CEGE's AVE is low (0.407), which suggests potential validity concerns. This could mean that the indicators for flow may not sufficiently capture the construct.

Table 9. Construct Reliability and Validity for flow and CEGE.

	Cronbach's Alpha (CA)	rho_A	Composite Reliability (CR)	Average Variance Extracted (AVE)
Threshold	≥0.7	≥0.7	≥0.7	≥0.5
CEGE	0.793	0.869	0.803	0.432
Flow	-	1.000	-	-

The SRMR (Standardized Root Mean Square Residual) value of 0.107 is very close to the maximum threshold of 0.10, suggesting the model is very close to being acceptable with room for improvement. d_ULS (Unweighted Least Squares Discrepancy) and d_G (Geodesic Discrepancy) lower values indicate a better fit. The Chi-Square value of 88.938 for the estimated model is at least equal to the saturated value; thus, we can count on the model fit. NFI = 0.641 < 0.80 indicates a poor fit. Low NFI could be due to sample size issues, 26 subjects (Table 10).

Table 10. Model Fit for Saturated and Estimated Models on flow and CEGE.

	Saturated Model	Estimated Model
SRMR	0.107	0.107
d_ULS	1.385	1.385
d_G	0.874	0.874
Chi-Square	88.938	88.938
NFI	0.641	0.641

6. Discussions

VR has emerged as a promising tool for STEM education, offering immersive experiences and active learning opportunities. In accordance with the specialized literature, our developed software meets a series of quality criteria such as usability, fidelity, and effectiveness, in accordance with the guidelines provided in the relevant literature [8,10,28].

An important limitation of this study is its reliance on a cross-sectional design with immediate post-intervention assessments. This prevents us from making empirical claims about the long-term educational impact of knowledge transfer.

This design was chosen to efficiently evaluate the immediate efficacy of the *Submarine Simulator.* However, the absence of longitudinal data limits our ability to assess whether observed improvements in underwater engineering and hydrodynamic principles persist over time or translate to other contexts. Future studies should employ longitudinal designs, with follow-up assessments at multiple intervals, to investigate the durability and transferability of these effects. Despite this limitation, the current findings provide valuable evidence of our VR tool's immediate benefits, laying the groundwork for more comprehensive evaluations.

The positive affordance questionnaire responses, where most students reported the software application as intuitive and supportive of task completion, suggest that the VR environment successfully bridged the gap between theoretical knowledge and simulated practice. Our VR application for submarine design and testing provides users with an

intuitive and engaging platform to explore complex engineering principles. The application attempts to represent hydrodynamic forces as accurately as possible for real-time computation, allowing for realistic experimentation and observation. This immersive environment fosters a deeper understanding of theoretical concepts by enabling direct manipulation and real-time feedback, transforming abstract equations into tangible outcomes, in accordance with the learning objectives [29].

Secondly, the data reveals that most participants experienced flow, with self-reported metrics (via the Flow State Scale questionnaire) indicating high levels of absorption and enjoyment. Notably, the strong correlation between affordance perceptions and flow suggests that the software's design elements, such as clear navigational cues and responsive interactions, facilitated this state by minimizing frustration and maximizing control.

This correlation extends to CEGE, including aspects like challenge, control, and feedback, which collectively enhanced immersion. Drawing from Csikszentmihalyi's flow theory, these results imply that VR's ability to create a "presence" in non-intuitive underwater settings through sensory realism and adaptive difficulty promotes flow more effectively than passive learning tools.

Furthermore, the application's design actively promotes problem-solving, critical thinking, and iterative design processes. These are essential conditions for an efficient learning process [6,34,41]. Users can prototype rapidly, test, and refine their submarine models, learning from failures and optimizing their designs in a risk-free virtual space.

The interdependencies among affordance, flow, and CEGE offer a compelling theoretical contribution, suggesting an expanded model for VR in education: affordances act as precursors to engagement, mediating flow through gamified structures. This builds on existing models like the Technology Acceptance Model (TAM), incorporating psychological flow as a key variable for non-intuitive domains.

Practically, these insights advocate for VR developers to prioritize iterative user testing focused on affordances to optimize flow and engagement, aligning with broader VR-in-education research that emphasizes user-centered design. This way, we can enhance immersion and learning outcomes in simulated environments. This approach could revolutionize underwater engineering pedagogy by transforming abstract concepts into immersive, hands-on simulations that mirror real-world challenges. Thereby, it fosters deeper conceptual understanding and develops practical skills through repeated virtual practice without physical risks. Additionally, it also bridges the gap between modern and traditional methods—such as lectures and static diagrams. Traditional methods often have difficulties conveying the dynamic complexities of underwater engineering. By incorporating VR into the learning environment, educators can support students in building the skills required to tackle innovative problem-solving in real-world scenarios.

These findings suggest to educators the potential for scalable VR integration, exemplified by hybrid curricula that blend immersive VR sessions with collaborative debriefing to strengthen overall student engagement throughout the learning process. Institutions could leverage these correlations to justify VR integration, emphasizing its role in preparing students for industry demands where simulation training is standard.

Our VR submarine application exemplifies the potential of immersive technologies to improve STEM education. We believe it empowers students to not only grasp complex scientific and engineering principles but also to innovate and apply their knowledge in a practical, meaningful way. By prioritizing active learning strategies, the original software design aligns with Johnson-Glenberg's (2019) call for a transformation in educational practices—one that centers on student interaction and engagement, which are crucial for effective STEM education [10].

Compared to complex Computer-Aided Design (CAD) software, which typically demands extensive training and significant time to master, our VR application's intuitive interface and immersive environment show promise for enabling users to quickly learn its building and testing capabilities. However, this observation is preliminary, as it does not include comparisons with other design tools or data on long-term skill retention and real-world application.

Additionally, the accessibility and cost of VR hardware pose significant challenges, particularly for educational institutions with limited budgets. High-quality VR systems often require substantial investment in equipment, maintenance, and technical support, creating disparities in access to these advanced learning tools. For instance, individual VR headsets like the Meta Quest 3 can cost between USD 500 and USD 600, while more advanced models such as the HTC Vive Pro range from USD 900 to USD 1200, not including additional expenses for VR-ready computers, laptops (USD 1200–2000), or classroom setup packs that may exceed reasonable prices for basic configurations. These financial challenges are further intensified by ongoing expenses for creating educational content and maintaining technical infrastructure, such as high-speed internet and powerful graphics cards for smooth VR experiences. These costs can be particularly burdensome for schools with limited budgets, deepening educational discrepancies and making it harder for all students around the globe to access cutting-edge learning tools.

Beyond financial barriers, VR adoption in education poses several practical challenges. Teachers and technical staff require specialized training to integrate VR effectively into lessons, as their role shifts from traditional instructors to facilitators that guide students through immersive, self-directed learning experiences. This transition demands not only technical proficiency with VR systems but also new pedagogical skills to manage interactive and dynamic virtual environments. Furthermore, overseeing an entire class of students using VR headsets can be particularly challenging for a single teacher. Unlike traditional classroom settings, where a teacher can easily monitor and engage with students, VR environments often isolate learners in individual headsets, making it difficult to maintain classroom discipline, provide real-time support, or ensure all students remain on task. This necessitates additional support staff, advanced classroom management tools, or smaller class sizes to ensure effective supervision and guidance, further complicating VR adoption in educational settings.

Compatibility issues, such as outdated computers or limited network capabilities, can also hinder adoption. Additionally, ensuring user comfort is key, which includes using ergonomic designs and session guidelines to reduce issues like motion sickness, making VR more accessible and enjoyable for students. Other barriers involve resistance to change among educators unfamiliar with VR, a shortage of high-quality educational content tailored to specific disciplines like underwater engineering, and potential health risks from prolonged use, such as eye strain, which can also slow the broader adoption of VR in educational environments.

Lastly, a potential barrier to adopting VR in education and training is technophobia, an irrational fear or aversion to new technologies that can cause anxiety, avoidance, or frustration among users. Practically, educators may experience heightened anxiety and low confidence during the initial use of VR, leading to resistance and limited implementation. Mitigation strategies, such as targeted training and gradual exposure, can build confidence and reduce barriers, enabling VR integration and adoption.

VR can enhance inclusion in education and training by creating immersive environments that accommodate diverse needs and overcome physical barriers. As noted in recent research, the current education landscape represents a unique opportunity to build accessible and inclusive virtual worlds from the very beginning, leveraging metaverse and

extended reality technologies for lifelong learning. This benefits students with disabilities through personalized simulations, adaptive training, and remote access, as evidenced by VR's effectiveness in special education [56].

Furthermore, VR promotes equitable learning pathways, such as STEM applications for underrepresented groups, fostering creativity and empathy while aligning with sustainable goals. By integrating supportive frameworks, educators can create inclusive ecosystems that reduce disparities and broaden participation.

7. Conclusions

The originality of our VR application is rooted in several key elements that differentiate it from existing educational tools. A cornerstone of this novelty is the customized physics engine specifically designed for the underwater simulation scene. Recognizing the limitations of generic physics engines in accurately representing complex hydrodynamics, we carefully re-engineered over 60% of the code. This extensive process delivered a better level of fidelity and realism to users when testing submarine models, far from simplistic games and into simulation territory. The realistic simulation of buoyancy, drag, and fluid dynamics allows for a truly authentic experience that attempts to mirror real-world underwater environments, a critical factor for effective engineering education.

Furthermore, the perceived quality and usability of the application have been independently validated through the affordance questionnaire, with results indicating a very high rating. This high rating underscores the intuitive design and ease of interaction within the virtual environment, ensuring that users can focus on learning and experimentation rather than struggling with the interface.

Perhaps the most compelling element of novelty is the absence of any other publicly available and mature application that offers a similar comprehensive platform for virtual submarine design and testing. While isolated simulations or design tools may exist, none integrate the full spectrum of design, testing, and realistic physics within an immersive VR environment to the degree our application does. This unique positioning fills a significant gap in STEM education tools, offering an unparalleled opportunity for students and professionals to engage with complex naval architecture and engineering principles in an accessible and engaging way.

These early findings indicate that the VR application's user-friendly design could reduce barriers for learners, potentially enabling a wider audience, including those unfamiliar with complex design software, to engage with advanced design and engineering concepts. This streamlined onboarding process appears to enhance engagement and optimize learning time, allowing users to transition quickly from learning the interface to experimenting and innovating. Further research, including comparative studies with tools like CAD and longitudinal assessments of skill durability, is needed to validate these observations and fully evaluate the application's educational impact.

Limits and Future Work

In the evaluation process of this software, we highlight an initial limitation related to the relatively small number of participants. Repeating the study with a larger sample size should be linked to accessing a group with similar characteristics. Further studies will focus on the relationship between the effectiveness of the training process and the characteristics of the participants involved in it.

Additionally, the small sample size and homogenous nature of the cohort, drawn from a single institution, can pose challenges to the generalizability of the findings. The uniformity in age, educational background, and institutional context may not reflect the

diversity of learners in other settings, such as different universities, engineering disciplines, or professional environments.

This homogeneity introduces potential sample bias, as the cohort's specific characteristics—such as their advanced engineering training or familiarity with technology—may have influenced their performance and engagement with the VR software, potentially inflating its perceived effectiveness. These limitations suggest that the study's results may not fully apply to broader populations, such as novice learners, students from varied academic backgrounds, or professionals in underwater engineering.

To address the limitations of the small and homogenous sample, further research should incorporate larger and more diverse participant groups, including students from multiple institutions, varying levels of expertise, and different demographic backgrounds. Such studies would enhance the generalizability of findings and reduce potential biases, providing a more robust understanding of the VR Submarine Simulator's effectiveness across diverse educational and professional contexts.

While the evaluation of VR software for educational purposes is comprehensive, it is important to recognize the inherent challenges and limitations in the process itself. One significant limitation lies in standardizing evaluation metrics across diverse VR applications. The immersive and interactive nature of VR introduces variables not typically found in traditional software, such as motion sickness and the difficulty in isolating the specific pedagogical gains directly attributable to the VR experience versus other learning modalities. Quantifying the subtle, yet powerful, influence of presence and immersion on learning outcomes remains a complex area of research.

There are some limitations concerning the statistical validity of certain measures. Specifically, AVE and discriminant validity indicators fell below standard thresholds, raising concerns about measurement precision and results. These issues may result from a small sample size, ambiguous questionnaire formulations, or language barriers. We acknowledge that the AVE values for the flow construct (0.407) and Core Elements of the Gaming Experience—CEGE (0.432) fall below the recommended threshold of 0.50 [57], indicating potential weaknesses in convergent validity.

Given the exploratory nature of this study and the novelty of the VR learning environment, we recognize the challenge in precisely defining latent constructs at this stage; thus, future iterations will involve item refinement and validation on a larger, homogenous sample.

While the current sample size of 26 is acceptable for exploratory PLS-SEM, we conclude that it constrains generalizability and statistical power; therefore, subsequent studies will involve larger samples to enable more robust validation, subgroup analyses, and improved reliability of parameter estimates.

To address another limitation in future studies, we recommend refining the affordance questionnaire items to more explicitly align with VR affordance and flow constructs. Clearer wording or adding illustrative examples to certain survey questions may improve participant comprehension and response accuracy, enhancing subsequent statistical robustness.

Another challenge involves the long-term retention and transfer of knowledge gained in VR environments. While initial studies often show strong immediate learning, more longitudinal research is needed to determine how well these skills and understandings translate to real-world applications and how durable they are over time.

Potential confounding factors further complicate VR implementation and research evaluation. These include variability in participants' prior VR and gaming experience, which can influence perceived immersion and learning outcomes; differences in hardware performance across institutions, leading to inconsistent user experiences; and environ-

mental factors like classroom setup, lighting, noise, or distractions that may diminish the sense of presence in VR simulations. In educational research contexts, such as studies on our *Submarine Simulator*, confounding often arises in media comparison designs where factors like instructional methods, novelty effects, or learner motivation are not adequately controlled, potentially skewing results on VR's efficacy compared to traditional methods. Addressing these requires standardized protocols for VR deployment, including guidelines for hardware setup, user onboarding processes to mitigate experience discrepancies, and consistent evaluation metrics tailored to VR's unique features, such as immersion scales and longitudinal tracking of skill transfer.

Future research should therefore focus on optimizing these environments to maximize their educational potential while addressing these challenges. This includes developing more robust and universally applicable evaluation frameworks that account for the unique characteristics of VR. Exploring how VR can be seamlessly integrated with other teaching methods to create blended learning experiences will be key to unlocking its full potential, ensuring that it complements, rather than replaces, traditional educational approaches.

Author Contributions: Conceptualization, A.-B.S., S.T. and R.-V.R.; methodology, A.-B.S., S.T. and R.-V.R.; software, A.-B.S.; validation, A.-B.S.; formal analysis, A.-B.S., S.T. and R.-V.R.; investigation, A.-B.S.; writing—original draft preparation, A.-B.S., S.T. and R.-V.R.; writing—review and editing, A.-B.S., S.T. and R.-V.R.; supervision, S.T. and R.-V.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data was created.

Acknowledgments: The research was run within doctoral studies and supervised by the coordinators from MINES Paris—PSL and the National University of Science and Technology "POLITEHNICA" Bucharest.

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

Appendix A

Affordance Questionnaire Items

On a scale from 1 to 5, where 1 is Strongly Disagree and 5 is Strongly Agree:

- I found the VR environment easy to navigate (using the menu, navigating closer/away from the submarine mode, changing modes, changing views, and everything related)
- The VR controls were easy to learn and use.
- The overall VR experience did NOT cause any discomfort or motion sickness.
- The VR environment provided clear visual indications of what I could interact with.
- The VR environment felt responsive to my actions.
- I felt confident about how I could interact with objects in the VR environment.
- At the end of the experience, I felt confident about how to build and test submarine models.
- I found it easy to personalize the appearance and functionality of my submarine in Build Mode.
- The Build Mode provided enough freedom for my creativity in designing a unique submarine model.
- The VR environment provided clear indications of what actions I could perform.
- The tools in VR felt intuitive to use.
- I found it easy to understand the function of the different tools available in VR.
- When I looked at two components, I could easily tell if they were snapped together.

- I believe that the VR simulation had a good representation of underwater fluid mechanics (in alignment and close to reality).
- I believe that all the represented forces and calculations during the underwater simulation were enough to understand the efficiency of my submarine model.
- I felt engaged and fully focused during the experiment.
- I felt comfortable in the virtual submarine simulation environment (scene 2).
- The experiment allowed me to iterate fast and efficiently over my submarine models.
- Overall—I believe that this learning experience in Virtual Reality brings real educational value in the process of prototyping a real, small-scale submarine.
- I did not feel that one hour had passed; I lost track of time.

Appendix B

Table A1. Commands and guidelines given to all participants.

Chapter 0: Video Introduction	• **The video tutorial contained an in-depth explanation of each function and mode available in the VR software application.**
Chapter 1: Building Mode	• Create a new submarine model. • Add a cube to the scene and move it to the center of the table. • Scale the cube proportionally. • Change the material of the cube to ABS. • Make sure the weight of the cube is between 7 and 10 kg. • Add a cylinder to the scene. • Snap the cylinder on top of the cube exactly in the middle using the "Face Snap" function. • Attach two motors to the cube, one motor on the left side of the cube, and one motor on the right side of the cube. • Save your model. • Move to simulation mode with the current model.
Chapter 2: Floatability Testing and Underwater navigation	• In front of you—on the virtual table you have two handles—use the left one to rotate around the object and the right one to zoom in and zoom out. • Below, under the table—you have two handles. Each one has the same color as the motor it controls. • Use the motors individually or together to move the ensemble underwater. • Experiment with navigating with the model underwater—rotate it/move forward/backwards. • Press "A" to change to a first-person view and use the joysticks to move underwater. • Press "A" again to return to the initial third-person view. • Restart the simulation using the green button on the console. • Exit the simulation using the red button on the console. • Open Q&A
Chapter 3: Exploratory Learning	• Create a model made of 4 shapes and 2 motors that sinks into the water between 1000 and 2000 m. The model needs to be as controllable as possible (can move forward, backwards, left, and right in a controlled manner). • Participants were informed that they had complete creative freedom to design their 3D submarine models.

Appendix C

Data and Data Distribution Review

Table A2. Distribution of data—affordance.

Statistic	Value
Count	26
Mean	85.69
Median	87.00
Std Deviation	6.92
Min	70.00
25%	81.25
75%	91.00
Max	98.00
Skewness	−0.42
Kurtosis	−0.42

Table A3. Distribution of data—flow.

Statistic	Value
Count	26
Mean	3.80
Median	3.92
Std Deviation	0.54
Min	2.50
25%	3.41
75%	4.19
Max	4.53
Skewness	−0.85
Kurtosis	0.21

Table A4. Distribution of data—CEGE.

Statistic	Value
Count	26
Mean	5.17
Median	5.24
Std Deviation	0.64
Min	3.39
25%	4.84
75%	5.66
Max	6.24
Skewness	−0.68
Kurtosis	0.58

All dimensions show left-skewed distributions, suggesting that while most users score moderately high, there are a few lower scores pulling the means down. AFF has the widest relative spread, while FSS and CEGE are more consistent. With only 26 points, these patterns should be interpreted cautiously; larger samples might reveal more about underlying normality or multimodality.

The following raw data has been utilized throughout the statistical analysis described in the section related to results:

Table A5. Raw data—questionnaire results for each participant in the study.

UserID	Affordance (AFF)	Flow (FSS)	Core Elements of Gaming Experience (CEGE)
290	73	3.778	4.303
709	83	4.028	5.455
227	81	4.528	6.152
512	93	4.222	5.212
177	91	4.389	5.788
443	90	4.194	5.788
187	76	3.972	4.667
159	89	4.333	5.152
830	93	4.194	5.667
670	88	3.778	4.909
590	79	3.556	4.333
595	83	3.361	4.818
682	82	3.972	4.818
341	87	4.361	5.515
411	91	3.861	5.273
840	91	3.194	5.788
388	94	3.972	5.061
723	94	4.111	5.273
985	84	2.528	4.939
779	79	3.361	3.394
298	80	3.333	4.303
484	98	4.528	6.242
770	87	3.222	5.273
955	87	3.722	5.697
642	70	2.500	4.939
256	85	3.778	5.636

References

1. Qorbani, H.S.; Arya, A.; Nowlan, N.S.; Abdinejad, M. ScienceVR: A Virtual Reality Framework for STEM Education, Simulation and Assessment. In Proceedings of the 2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), Taichung, China, 15–17 November 2021; pp. 267–275. [CrossRef]
2. Ward, T.; Ortega-Moody, J.; Khoury, S.; Wheatley, M.; Jenab, K. Virtual reality platforms for K-12 STEM education. *Manag. Sci. Lett.* **2025**, *15*, 193–204. [CrossRef]
3. Smith, J.R.; Snapp, B.; Madar, S.; Brown, J.R.; Fowler, J.; Andersen, M.; Porter, C.D.; Orban, C. A Smartphone-Based Virtual Reality Plotting System for STEM Education. *PRIMUS* **2023**, *33*, 1–15. [CrossRef]
4. Nersesian, E.; Vinnikov, M.; Lee, M.J. Travel Kinematics in Virtual Reality Increases Learning Efficiency. In Proceedings of the 2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), St. Louis, MO, USA, 10–13 October 2021; pp. 1–5. [CrossRef]
5. Ha, O. Development of a low-cost immersive virtual reality solution for STEM classroom instruction: A case in Engineering Statics. In Proceedings of the 2020 2nd International Workshop on Artificial Intelligence and Education (WAIE 2020), Montreal, QC, Canada, 6–8 November 2020; Association for Computing Machinery: New York, NY, USA; pp. 64–68. [CrossRef]
6. Janonis, A.; Kiudys, E.; Girdžiūna, M.; Blažauskas, T.; Paulauskas, L.; Andrejevas, A. *Escape the Lab: Chemical Experiments in Virtual Reality*; Springer: Cham, Switzerland, 2020; pp. 273–282. [CrossRef]
7. Laseinde, O.T.; Dada, D. Enhancing teaching and learning in STEM Labs: The development of an android-based virtual reality platform. *Mater. Today Proc.* **2023**, *105*, 240–246. [CrossRef]
8. Roy, S.; Chang, D.Y.; Stocksdale, G.; Gunasekera, M.; Rizwan-uddin, R. Recent Advances in VR Labs for Use in STEM Education. In Proceedings of the Polytechnic University of Valencia Congress, Tenth International Conference on Higher Education Advances, Valencia, Spain, 18–21 June 2024. [CrossRef]
9. Klän, W. Using Virtual Reality to Improve STEM Education. Ph.D. Thesis, Carleton University, Ottawa, Canada, 2023. [CrossRef]
10. Johnson-Glenberg, M.C. *The Necessary Nine: Design Principles for Embodied VR and Active Stem Education*; Springer: Singapore, 2019; pp. 83–112. [CrossRef]

11. Semerikov, S.; Mintii, M.; Mintii, I. Review of the course "Development of Virtual and Augmented Reality Software" for STEM teachers: Implementation results and improvement potentials. In Proceedings of the International Workshop on Augmented Reality in Education, Kryvyi Rih, Ukraine, 11 May 2021. [CrossRef]
12. Pirker, J.; Dengel, A.R.; Holly, M.S.; Safikhani, S. Virtual Reality in Computer Science Education: A Systematic Review. In Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology, Ottawa, Canada, 1–4 November 2020. [CrossRef]
13. Mystakidis, S.; Fragkaki, M.; Filippousis, G. Ready Teacher One: Virtual and Augmented Reality Online Professional Development for K-12 School Teachers. *Computers* **2021**, *10*, 134. [CrossRef]
14. Standford University. Stanford Researchers Release Virtual Reality Simulation that Transports Users to Ocean of the Future. 2019. Available online: https://news.stanford.edu/stories/2016/10/virtual-reality-simulation-transports-users-ocean-future (accessed on 21 August 2025).
15. Purdue University Northwest. CIVS Developed Underwater VR Lab Modules for An NSF-Funded STEM Project. 2023. Available online: https://www.pnw.edu/civs/2023/11/07/civs-technology-applied-to-nsf-funded-national-stem-project/ (accessed on 21 August 2025).
16. Fauville, G.; Voski, A.; Mado, M.; Lantz-Andersson, A. Underwater virtual reality for marine education and ocean literacy: Technological and psychological potentials. *Environ. Educ. Res.* **2024**, 1–25. [CrossRef]
17. Aldhaheri, S.; Hu, Y.; Xie, Y.; Wu, P.; Kanoulas, D.; Liu, Y. Underwater Robotic Simulators Review for Autonomous System Development. *arXiv* **2025**, arXiv:2504.06245v1.
18. Rosette, M.; Hannah Kolano, H.; Holm, C.; Hollinger, G.A.; Marburg, A.; Pickett, M.; Davidson, J.R. WAVE: An Open-Source Under Water Arm-Vehicle Emulator. Available online: https://research.engr.oregonstate.edu/rdml/sites/research.engr.oregonstate.edu.rdml/files/icra24_1333_fi.pdf (accessed on 21 August 2025).
19. Yoon, J.; Son, B.; Lee, D. Comparative Study of Physics Engines for Robot Simulation with Mechanical Interaction. *Appl. Sci.* **2023**, *13*, 680. [CrossRef]
20. Sinnott, C.; Liu, J.; Matera, C.; Halow, S.; Jones, A.; Moroz, M.; Mulligan, J.; Crognale, M.; Folmer, E.; MacNeilage, P. Underwater Virtual Reality System for Neutral Buoyancy Training: Development and Evaluation. In Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology (VRST '19), Parramatta, Australia, 12–15 November 2019; Association for Computing Machinery: New York, NY, USA; pp. 1–9. [CrossRef]
21. Haklidir, M.; Guven, A.F.; Eroglu, O.; Aldogan, D.; Tasdelen, I. High Fidelity Modelling and Simulation of Submarine in a Commercial Computer Generated Forces Toolkit. In Proceedings of the Society for Modeling & Simulation International [SCS] 2009 Summer Computer Simulation Conference (SCSC'09), Istanbul, Turkey, 13–16 July 2009. Available online: https://www.researchgate.net/publication/234715910_High_Fidelity_Modeling_and_Simulation_of_Submarine_in_a_Commercial_Computer_Generated_Forces_Toolkit (accessed on 21 August 2025).
22. Chin, C.S.; Kamsani, N.B.; Zhong, X.; Cui, R.; Yang, C. Unity3D Serious Game Engine for High Fidelity Virtual Reality Training of Remotely-Operated Vehicle Pilot. In Proceedings of the 10th International Conference on Modelling, Identification and Control (ICMIC), Guiyang, China, 2–4 July 2018; pp. 1–6. [CrossRef]
23. Tohoku University. "Let it Flow: Recreating Water Flow for Virtual Reality." ScienceDaily. ScienceDaily, 21 September 2023. Available online: www.sciencedaily.com/releases/2023/09/230920111215.htm (accessed on 21 August 2025).
24. Makransky, G.; Terkildsen, T.S.; Mayer, R.E. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learn. Instr.* **2019**, *60*, 225–236. [CrossRef]
25. Huang, F.; Chen, X.; Xu, Y.; Yang, X.; Chen, Z. Immersive virtual simulation system design for the guidance, navigation and control of unmanned surface vehicles. *Ocean. Eng.* **2023**, *281*, 114884. [CrossRef]
26. Deck, C. Virtual reality and science, technology, engineering, and mathematics education. *Elsevier Ebooks* **2023**, 189–197. [CrossRef]
27. Paskova, A. Features of application of immersive technologies of virtual and augmented reality in higher education. *Vestn. Majkopskogo Gos. Tehnol. Univ.* **2022**, 14. [CrossRef]
28. Zikas, P.; Papagiannakis, G.; Lydatakis, N.; Kateros, S.; Ntoa, S.; Adami, I.; Stephanidis, C. Immersive visual scripting based on VR software design patterns for experiential training. *Vis. Comput.* **2020**, *36*, 1965–1977. [CrossRef]
29. Matovu, H. Immersive virtual reality for science learning: Design, implementation, and evaluation. *Stud. Sci. Educ.* **2022**, *59*, 205–244. [CrossRef]
30. Danmali, S.S.; Onansanya, S.A.; Atanda, F.A.; Abdullahi, A. Application of Virtual Reality in STEM Education for Enhancing Immersive Learning and Performance of At-Risk Secondary School Students. *Int. J. Res. Innov. Soc. Sci.* **2024**, *8*, 3971–3984. [CrossRef]
31. Li, S. An Analysis of Immersive Communication in VR Documentary: A Case Study of Planet Earth II. *Commun. Humanit. Res.* **2023**, *16*, 110–115. [CrossRef]
32. Zhu, J. Immersive Experience Design of Digital Media Interactive Art Based on Virtual Reality. *Comput. Aided Des. Appl.* **2024**, *21*, 164–177. [CrossRef]

33. Harris, T.M. Immersive virtual reality and spatial analysis, Chapters. In *Handbook of Spatial Analysis in the Social Sciences*; Rey, S.J., Franklin, R.S., Eds.; Edward Elgar Publishing: Cheltenham, UK, 2022; Chapter 20; pp. 336–351. Available online: https://ideas.repec.org/h/elg/eechap/19110_20.html (accessed on 21 August 2025).
34. Azofeifa, J.D.; Rueda-Castro, V.; Gonzalez-Gomez, L.J.; Noguez, J.; Caratozzolo, P. Revolutionizing Training in Continuing Engineering Education: An Experimental Proposal Using the Multi-Sensory Virtual Decision-Making Center. *IEEE Access* **2024**, *12*, 192440–192448. [CrossRef]
35. Truchly, P.; Medvecky, M.; Podhradsky, P.; Vanco, M.P. Virtual Reality Applications in STEM Education. In Proceedings of the 2018 16th International Conference on Emerging eLearning Technologies and Applications (ICETA), Stary Smokovec, Slovakia, 15–16 November 2018; pp. 597–602. [CrossRef]
36. Ma, T.; Xiao, X.; Wee, W.G.; Han, C.Y.; Zhou, X. A 3D Virtual Learning System for STEM Education. In *Virtual, Augmented and Mixed Reality*; Applications of Virtual and Augmented Reality; VAMR 2014; Lecture Notes in Computer Science; Shumaker, R., Lackey, S., Eds.; Springer: Cham, Switzerland, 2014; Volume 8526. [CrossRef]
37. Drossis, G.; Birliraki, C.; Stephanidis, C. Interaction with Immersive Cultural Heritage Environments Using Virtual Reality Technologies. In *HCI International 2018–Posters' Extended Abstracts*; HCI 2018; Communications in Computer and Information Science; Stephanidis, C., Ed.; Springer: Cham, Switzerland, 2018; Volume 852, pp. 177–183. [CrossRef]
38. Gómez, P.M.M.; Gómez, R.J.M.; Antonio, D.; Borré, F. Virtual Immersion in STEM Education: A MICMAC Study on How Virtual Reality Impacts the Understanding and Application of Scientific Concepts. *Evol. Stud. Imaginative Cult.* **2024**, *8*, 162–173. [CrossRef]
39. Johnson, D.; Mamani, B.; Salas, C. CollabVR: VR Testing for Increasing Social Interaction between College Students. *Computers* **2024**, *13*, 40. [CrossRef]
40. Lytras, M.D.; Papadopoulou, P.; Misseyanni, A.; Marouli, C.; Alhalabi, W.; Daniela, L. Moving virtual and augmented reality in the learning cloud: Design principles for an agora of active visual learning services in stem education. In Proceedings of the 9th International Conference on Education and New Learning Technologies, Barcelona, Spain, 3–5 July 2017; pp. 7673–7678. [CrossRef]
41. Mo, J. Improving the Middle School STEM Education in Rural China Through Virtual Reality. In Proceedings of the 4th International Conference on Educational Reform, Management Science and Sociology (ERMSS 2023), Detroit, MI, USA, 7–8 January 2023; Volume 9. [CrossRef]
42. Lynch, T.; Ghergulescu, I. Review of virtual labs as the emerging technologies for teaching stem subjects. In Proceedings of the 11th International Technology, Education and Development Conference, Valencia, Spain, 6–8 March 2017; pp. 6082–6091. [CrossRef]
43. Sun, W.; Chen, Q. The Design, Implementation and Evaluation of Gamified Immersive Virtual Reality (IVR) for Learning: A Review of Empirical Studies. In Proceedings of the 17th European Conference on Games Based Learning, Enschede, The Netherlands, 5–6 October 2023. [CrossRef]
44. Gerini, L.; Delzanno, G.; Guerrini, G.; Solari, F.; Chessa, M. Gamified Virtual Reality for Computational Thinking. In Proceedings of the Gamify 2023: Proceedings of the 2nd International Workshop on Gamification in Software Development, Verification, and Validation, San Francisco, CA, USA, 4 December 2023; pp. 13–21. [CrossRef]
45. Ruscanu, A.-M.; Ciupe, A.; Meza, S. arPcTECHture–a gamified educational 3D virtual world for introductory concepts in computer architecture. In Proceedings of the IEEE Global Engineering Education Conference, Tunis, Tunisia, 28–31 March 2022; pp. 1437–1442. [CrossRef]
46. Tiefenbacher, F. Evaluation of Gamification Elements in a VR Application for Higher Education. In *Systems, Software and Services Process Improvement*; EuroSPI 2020; Communic Ations in Computer and Information Science; Yilmaz, M., Niemann, J., Clarke, P., Messnarz, R., Eds.; Springer: Cham, Switzerland, 2020; Volume 1251. [CrossRef]
47. Gibson, J.J. The Theory of Affordance. In *Ecological Approach to Visual Perception*; Houghton Mifflin Company: Boston, MA, USA, 1979. Available online: https://monoskop.org/images/c/c6/Gibson_James_J_1977_1979_The_Theory_of_Affordances.pdf (accessed on 21 August 2025).
48. Jackson, S.A.; Marsh, H. Development and validation of a scale to measure optimal experience: The Flow State Scale. *J. Sport Exerc. Psychol.* **1996**, *18*, 17–35. [CrossRef]
49. Calvillo-Gámez, E.H.; Cairns, P.; Cox, A.L. Assessing the Core Elements of the Gaming Experience. In *Evaluating User Experience in Games*; Human-Computer Interaction Series; Bernhaupt, R., Ed.; Springer: London, UK, 2010. [CrossRef]
50. AlGerafi, M.A.M.; Zhou, Y.; Oubibi, M.; Wijaya, T.T. Unlocking the Potential: A Comprehensive Evaluation of Augmented Reality and Virtual Reality in Education. *Electronics* **2023**, *12*, 3953. [CrossRef]
51. Bernhard, K.; Henning, S. Development of a software-tool to evaluate the tolerability of different VR-movement types. *Health Technol.* **2024**, *14*, 781–790. [CrossRef]
52. Ringle, C.M.; Wende, S.; Becker, J.-M. "SmartPLS3" Boenningstedt: SmartPLS GmbH. 2015. Available online: http://www.smartpls.com (accessed on 21 August 2025).

53. Hair, J.F.; Hult, G.T.M.; Ringle, C.M.; Sarstedt, M. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 2nd ed.; Sage Publications: Thousand Oaks, CA, USA, 2017.
54. Henseler, J.; Ringle, C.M.; Sinkovics, R.R. The use of partial least squares path modeling in international marketing. *Adv. Int. Mark.* **2009**, *20*, 277–319. [CrossRef]
55. Cohen, J. *Statistical Power Analysis for Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988. Available online: https://utstat.utoronto.ca/brunner/oldclass/378f16/readings/CohenPower.pdf (accessed on 21 August 2025).
56. Anastasovitis, E.; Roumeliotis, M. Enhanced inclusion and accessibility in education and training through virtual worlds. *Metaverse* **2024**, *5*, 2836. [CrossRef]
57. Hair, J.F.; Risher, J.J.; Sarstedt, M.; Ringle, C.M. When to use and how to report the results of PLS-SEM. *Eur. Bus. Rev.* **2019**, *31*, 2–24. [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.

computers

MDPI

Article

Examining the Flow Dynamics of Artificial Intelligence in Real-Time Classroom Applications

Zoltán Szűts [1,*], Tünde Lengyelné Molnár [1], Réka Racskó [1], Geoffrey Vaughan [1,*], Szabolcs Ceglédi [2] and Dalma Lilla Dominek [2]

[1] Doctoral School of Education, Faculty of Pedagogy, Eszterházy Károly Catholic University, Eszterházy Square 1, 3300 Eger, Hungary; lengyelne.tunde@uni-eszterhazy.hu (T.L.M.); racsko.reka@uni-eszterhazy.hu (R.R.)
[2] Department of Social Communication, Ludovika University of Public Service, 1083 Budapest, Hungary; cegledi.szabolcs@uni-nke.hu (S.C.); dominek.dalma.lilla@uni-nke.hu (D.L.D.)
* Correspondence: szuts.zoltan@uni-eszterhazy.hu (Z.S.); geoffrey.vaughan@uni-eszterhazy.hu (G.V.)

Abstract: The integration of artificial intelligence (AI) into educational environments is fundamentally transforming the learning process, raising new questions regarding student engagement and motivation. This empirical study investigates the relationship between AI-based learning support and the experience of flow, defined as the optimal state of deep attention and intrinsic motivation, among university students. Building on Csíkszentmihályi's flow theory and current models of technology-enhanced learning, we applied a validated, purposefully developed AI questionnaire (AIFLQ) to 142 students from two Hungarian universities: the Ludovika University of Public Service and Eszterházy Károly Catholic University. The participants used generative AI tools (e.g., ChatGPT 4, SUNO) during their academic tasks. Based on the results of the Mann–Whitney U test, significant differences were found between students from the two universities in the immersion and balance factors, as well as in the overall flow score, while the AI-related factor showed no statistically significant differences. The sustainability of the flow experience appears to be linked more to pedagogical methodological factors than to institutional ones, highlighting the importance of instructional support in fostering optimal learning experiences. Demographic variables also influenced the flow experience. In gender comparisons, female students showed significantly higher values for the immersion factor. According to the Kruskal–Wallis test, educational attainment also affected the flow experience, with students holding higher education degrees achieving higher flow scores. Our findings suggest that through the conscious design of AI tools and learning environments, taking into account instructional support and learner characteristics, it is possible to promote the development of optimal learning states. This research provides empirical evidence at the intersection of AI and motivational psychology, contributing to both domestic and international discourse in educational psychology and digital pedagogy.

Keywords: artificial intelligence (AI); flow experience; higher education; digital pedagogy; generative AI

1. Introduction

The rise of artificial intelligence (AI) in education is one of the most significant paradigm shifts of the last decade. There are different types of AI, such as machine learning, natural language processing, and generative models (e.g., ChatGPT and Gemini), that are becoming increasingly prevalent in learning support. These tools aim to automate

the teaching and learning process, but also to enable a deeper understanding of, and support for, learner motivation, engagement, and the learning experience. A key aspect of optimizing the learning experience is the flow experience. The presence of this experience correlates with learner performance, satisfaction, and intrinsic motivation [1]. This paper aims to review the relationship between AI tools and the flow experience, particularly in light of empirical research conducted among university students.

According to flow theory, an optimal experience occurs when an individual's abilities are balanced with the challenges of the task, and when goals are clear and the individual receives immediate feedback [1]. In an educational setting, this state results in high levels of engagement and joyful learning experience and deep processing [2]. For university students, flow plays a particularly important role in self-directed learning, where autonomy and intrinsic motivation are the determining factors [3]. In digital learning, the flow experience is influenced by a number of factors, such as the use of tools, instructor presence, the nature of the course material, and the degree of interactivity [4]. Ref. [5] was among the first to point out that computer-assisted learning can result in a flow state when learners feel relevant and challenged by the tasks. The Hungarian educational psychology literature also addresses the factors that trigger flow. Another article [6] emphasize that the quality of the learning experience is significantly influenced by the pedagogical environment and the feedback provided during the learning process.

The integration of AI tools into education can have three main impacts: (1) Providing personalized learning pathways, (2) providing real-time feedback, and (3) analyzing learner data for learning analytics [7]. Systems such as intelligent tutoring systems (ITS), recommender systems, and predictive models can be particularly useful in higher education, where students' learning styles and prior knowledge can vary significantly [8]. Generative AI tools, such as ChatGPT, Copilot, and Gemini, have opened up a new dimension in learning support. These systems can interact with users, providing information, answering questions, offering feedback, and encouraging reflection; answer questions; provide feedback; and encourage reflective thinking [9].

The study of the flow experience is becoming increasingly important in the context of AI-based learning environments. To this end, [10] developed the Artificial Intelligence Factor-Facilitated Flow Questionnaire (AIFLQ), which, in addition to the classical flow dimensions (immersion, challenge–skill balance, clear goals, and immediate feedback), measures an AI-specific factor. This new factor focuses on elements such as technological reliability, personalization, and learner autonomy when using generative AI tools. The AIFLQ includes statements such as 'While using the AI tools, I felt that the technology supported my individual learning needs' and 'Using generative AI tools increased my motivation to learn'. These elements help us understand how AI contributes to the flow experience. The AIFLQ has been validated among university students and has high internal reliability (Cronbach's $\alpha > 0.80$). Its multidimensional factor structure corresponds to the integrated framework of flow theory and instructional technology models [10].

Several studies have investigated the relationship between AI tools and the flow experience. For example, [11] used an intelligent tutoring system in a university programming course and found that the system's real-time feedback and adaptivity significantly increased students' flow experience. Similar findings emerged from the research of [12], who observed heightened immersion and a greater sense of competence during language learning tasks facilitated by generative AI tools. Others [13] examined the impact of generative AI tools (e.g., ChatGPT) on flow among a Chinese university population. Their findings revealed that female students and those with higher self-efficacy exhibited higher flow scores. Positive attitudes and experiences with AI tools also contributed to deeper engagement.

Although AI tools can optimize the technological aspects of the learning environment, the pedagogical framework and instructor support are key to developing a positive learning experience [14]. Structured implementation, clear articulation of learning objectives and a reflective learning culture can all multiply the effectiveness of AI tools. Studies [15] have shown that, in university courses where instructors facilitated the use of AI tools, students reported a significantly higher flow experience than in courses where the introduction of tools was left to self-discovery. The development of the flow experience is influenced not only by technology and pedagogy, but also by background variables of the learner. International research [6,12,13,16] suggests that gender, educational attainment, experience with AI tools, and self-efficacy can all significantly impact flow intensity. Females more often reported higher levels of immersion, while students with less experience of technology had lower levels of challenge and confidence in AI. Psychological factors such as learning orientation, grit, and self-regulation ability can also predict learners' ability to achieve a state of flow when using AI tools [17]. Grit is defined as an individual's persistence and commitment to long-term goals, while self-regulation is the ability to consciously manage, plan, and control the learning process [18,19]. These qualities enable learners to adapt more effectively to challenges and have more immersive learning experiences.

A review of the existing literature shows that AI tools can have a positive impact on university students' flow experience when embedded in an appropriate pedagogical framework. The introduction of the AIFLQ questionnaire offers new ways to measure the combined effects of flow and AI experience, supporting future empirical research in this area. However, it is important to emphasize that technological innovation alone does not guarantee the achievement of a flow state. A deep, immersive learning experience requires the combination of instructor mediation, learner psychological characteristics and environmental factors. Further research is needed to explore the long-term effects of AI tools and identify the needs of students from different disciplines.

Although previous studies have examined the relationship between artificial intelligence (AI) tools and flow in learning environments, they have typically focused on specific use cases or lacked a comprehensive framework for evaluating the interaction between AI-supported learning and demographic and psychological variables. In particular, little is known about the performance of validated flow instruments, such as the Artificial Intelligence-Facilitated Flow Questionnaire (AIFLQ), across diverse learner profiles in real-time academic contexts. This study addresses these gaps by examining how gender, educational attainment, and personality type shape the flow experience in higher education settings. This contributes new empirical insights to educational psychology and digital pedagogy.

2. Materials and Methods

The integration of artificial intelligence tools in higher education is a rapidly evolving field. It offers opportunities to enhance learner motivation and engagement, as well as presenting complex challenges that deepen our understanding of the learning process [7,9]. While the concept of the 'flow experience' as an optimal learning state has been studied extensively [1], the specific effects of AI-supported learning environments in this regard remain under-explored [12]. In this article, the authors use the term "artificial intelligence-assisted" to point out that a given process, activity, or system relies in some way on artificial intelligence (AI) tools, algorithms, or services to improve its operation or efficiency. Here, AI is used to personalize and support the learning experience for the purpose of content creation. Current research indicates that AI tools not only serve as instruments, but also function as active learning support systems, offering immediate feedback, adaptability, and personalized learning pathways [8,11]. While these features have the potential to facilitate

the development of the flow state, the mediating or moderating role of individual factors such as gender, personality traits, and prior learning background in this context has not yet been adequately analyzed [13,18].

It is particularly important to explore the role of psychological personality variables such as self-regulation, grit, and motivational profiles, as these have a profound impact on learners' flow experience and interaction with technology [17,18]. Previous findings suggest that demographic and psychological factors can significantly impact the effectiveness of AI-based learning environments, but further empirical studies are needed to explore the exact mechanisms and relationships [12,13].

This study aims to address this research gap by investigating the effects of AI on flow status, focusing particularly on gender differences, prior educational attainment and personality traits, which may play a moderating role in developing optimal learning experiences.

Several consistent patterns and gaps have emerged from the reviewed literature that inform the present hypotheses. While prior studies suggest that AI tools may enhance learner engagement and immersion [11,12], they often fail to consider individual learner characteristics, such as gender or educational background [6,13]. Research also suggests that women may experience greater emotional engagement and that educational experience can impact self-regulated learning strategies, both of which are important for achieving flow [18,19]. However, there is a lack of conclusive evidence on whether such demographic and psychological variables systematically influence the flow state in AI-supported learning. Based on this body of evidence and the conceptual frameworks outlined earlier, the following four hypotheses have been developed.

2.1. Hypothesis

The integration of artificial intelligence (AI) tools into higher education presents both opportunities and challenges for students and educators alike [20]. While AI-based learning support has the potential to enhance engagement and motivation, understanding how it affects students' deep focus and optimal learning experiences remains a complex task. Achieving and sustaining a state of flow—a psychological state characterized by full immersion and intrinsic motivation—is already difficult in traditional learning environments, and the introduction of AI adds new variables to this dynamic. Moreover, individual differences such as gender, educational background, and personality traits may influence how students experience flow during AI-supported learning. Previous research highlights that technology's effectiveness in education often depends not only on access but also on learners' characteristics and contextual factors. However, the role of AI tools in fostering flow has been understudied, and little is known about whether demographic and personality variables moderate this experience. This study aims to fill this gap by examining four specific hypotheses related to AI usage, gender differences, educational qualifications, and personality types in relation to flow among university students. By doing so, it contributes to a deeper understanding of how AI can be consciously designed and implemented to support optimal learning states.

Therefore, as mentioned above, based on the literature on the topic and experience, the authors have established four hypotheses during research efforts.

H1. *The use of AI-based learning support is positively associated with the experience of flow among university students.*

H2. *Female students report significantly higher scores on the immersion factor of flow compared to male students.*

H3. *Students with higher educational qualifications demonstrate significantly higher overall flow scores than those with lower qualifications.*

H4. *Personality type has no significant effect on the experience of flow among university students.*

2.2. Methods

The primary aim of this study was to examine the relationship between background variables—namely, gender, educational attainment, institution attended, and personality traits—and the experience of flow during AI-supported learning environments. Specifically, we sought to understand how these factors influence students' engagement and optimal learning states while using AI tools in academic settings.

To test the proposed hypotheses, a quantitative research design was employed. Data collection took place during regular classes at two Hungarian universities: Ludovika University of Public Service (LUPS) and Eszterházy Károly Catholic University (EKCU). This approach allowed for capturing authentic student experiences in real-world AI-enhanced educational contexts.

Statistical analyses were conducted using non-parametric tests suitable for the flow measurement variables. The Kruskal–Wallis test was applied to investigate differences across multiple groups, such as varying levels of educational attainment, while the Mann–Whitney U test was used to assess pairwise comparisons, for example between genders or institutions. Descriptive statistics were also calculated to summarize the results of the various flow factors, providing an overview of the main tendencies and variability within the sample.

Together, these methods enabled a rigorous examination of how demographic and personality variables relate to the flow experience in AI-supported learning, contributing valuable insights to the intersection of educational psychology and digital pedagogy.

The data were collected using a purposefully developed and validated questionnaire tailored to assess the experience of flow within AI-supported learning environments. A total of 142 students participated in the study, with 68 respondents from Eszterházy Károly Catholic University (EKCU) and 74 from Ludovika University of Public Service (LUPS) (Figure 1). During the test, students filled out a questionnaire accessible via QR code during class. Completion was anonymous and voluntary, and no prior information was provided. We did not explain the purpose of the test because we were interested in their opinions and did not want to influence them. Pie and bar charts were employed to visually represent demographic distributions and descriptive statistics in a clear and accessible manner. Pie charts were used exclusively for nominal categorical variables (e.g., gender, academic orientation), where proportions are easily interpreted. For continuous variables (e.g., immersion, balance, AI factor), bar charts displaying means and standard deviations were chosen to illustrate central tendency and variability. These visual formats are appropriate given the nature of the data and are commonly used in psychological and social science reporting to enhance interpretability for a broad readership.

Due to the ordinal nature of the flow measurement scales and the non-normal distribution of the data, as confirmed by preliminary normality tests, non-parametric statistical methods were employed to test the proposed hypotheses. The Mann–Whitney U test was used to compare two independent groups (e.g., gender or education level), and the Kruskal–Wallis test was used to compare more than two categories (e.g., age groups). The threshold for statistical significance was set at $p < 0.05$. Where multiple comparisons were conducted, Bonferroni corrections were applied to control for Type I error inflation. These tests were selected to align with the constraints of the dataset in terms of the measurement scale and sample size.

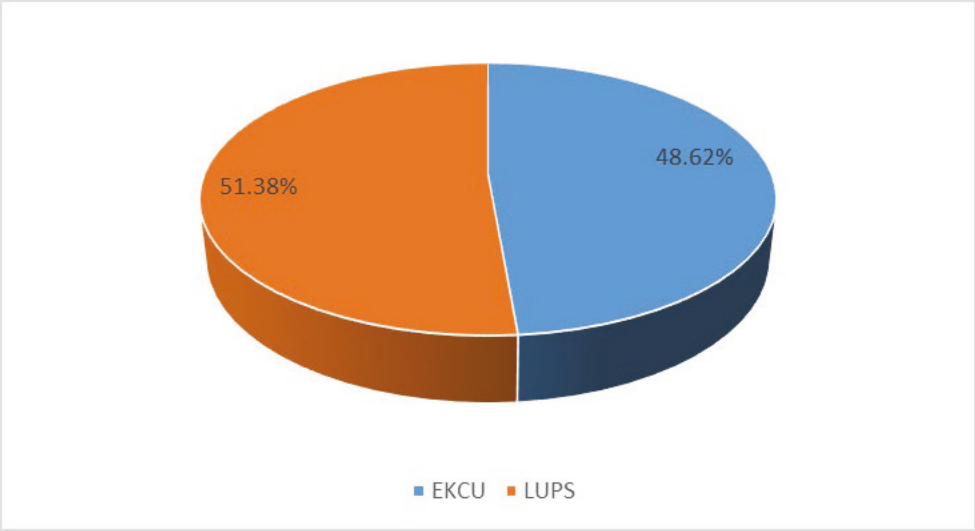


Figure 1. Institution of responders.

Given that the data in several instances deviated from normality and group variances were unequal, the assumptions of parametric tests were not met. Accordingly, non-parametric methods were employed as they provide more robust and reliable results under such conditions.

In the sample, approximately one-third of the participants were male, while two-thirds were female. This distribution offers a meaningful lens through which to consider the dynamics of flow in AI-supported learning, as gender-related factors may influence how learners engage with technology-enhanced educational environments (Figure 2).

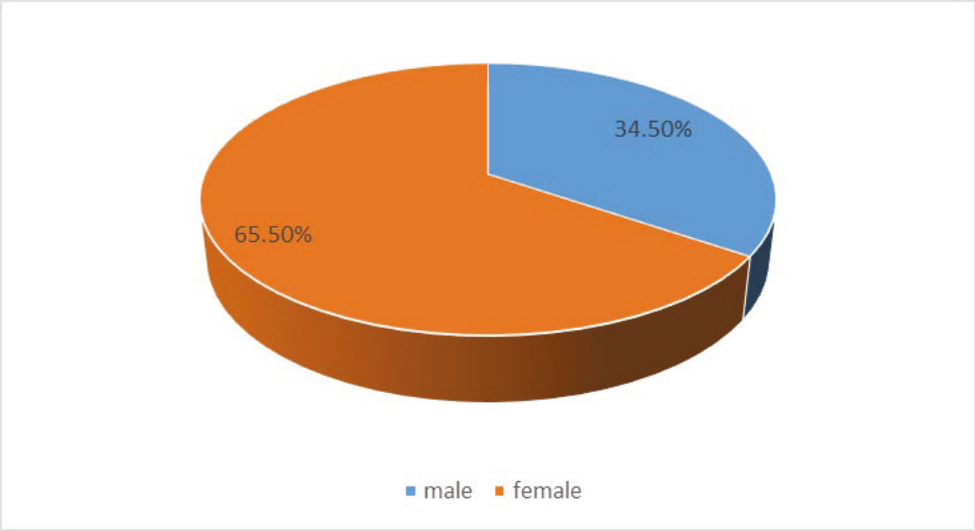


Figure 2. Gender of responders.

The participants represented a diverse age range, with the majority falling into the 18–25 age group (52.1%). This was followed by 26–33 years (16.2%), 41–50 years (12.7%), 34–40 years (14.1%), and 51 years or older (4.9%). This varied age composition suggests that AI-supported learning environments are engaging for learners across different life stages, offering valuable insights into how flow is experienced in both traditional and non-traditional student populations (Table 1).

Table 1. Age of responders.

Age Groups					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	18–25	53	52.1	52.1	47.3
	26–33	20	16.2	16.2	65.2
	34–40	13	12.7	12.7	76.8
	41–50	19	14.1	14.1	93.8
	51+	7	4.9	4.9	100.0
	Total	142	100.0	100.0	
Missing		0	0.0		
Total		142	100.0		

Participants were also asked to identify the academic field that best reflects their personality or way of thinking. The majority (46, 5.8%) associated themselves with Public Administration and Legal Studies, followed by Humanities and Social Sciences (43%), and STEM (science, technology, engineering, and mathematics) and Environmental Sciences (10.6%). This distribution suggests a strong representation of individuals who are oriented toward human-centered, interpretative, and socially engaged domains, while also including those with analytical and systematic thinking styles that align with legal systems and complex regulatory frameworks. Such diversity may influence how learners relate to AI-supported learning environments and experience flow in different cognitive and disciplinary contexts (Figure 3).

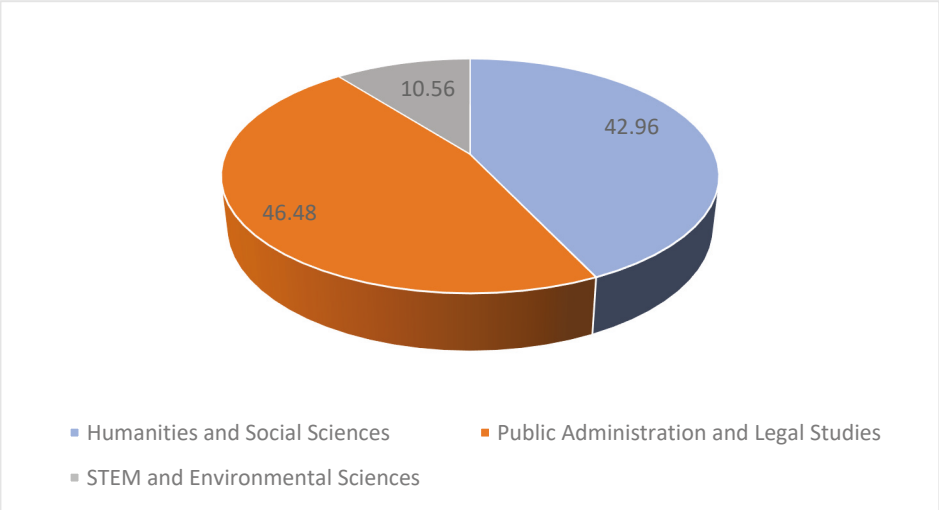


Figure 3. Personality traits of responders.

The questionnaire consisted of 24 items, systematically divided into three distinct flow-related factors, each represented by eight questions. These factors were: immersion, which captures the degree of deep engagement and absorption in learning activities; balance, reflecting the perceived equilibrium between challenge and skill; and the AI factor, which specifically addresses the influence and role of AI tools in the learning process.

To assess the participants' flow experience during the task, we used the AI and Flow Learning Questionnaire (AIFLQ) developed by [10]. This instrument is designed to measure the flow state in learning environments that involve artificial intelligence. It captures flow through a combination of subjective self-report and objective indicators recorded on a visual flow measurement board.

The AIFLQ focuses on three core dimensions of the flow experience: Balance, Immersion, and AI Support. Each of these dimensions is rated on a scale from 0 to 100%, and together they form the basis for calculating the Total Flow Score.

- Balance reflects the perceived match between the task's difficulty and the individual's skill level.
- Immersion indicates how deeply the participant is engaged and mentally involved in the task.
- AI evaluates the extent to which the AI system contributes positively to maintaining the user's flow.

The Total Flow Score is computed as the average of the three dimensions. A score above 80% is considered indicative of a sustained or optimal flow state, suggesting a high level of engagement and learning effectiveness during the activity.

3. Results

The final sample comprised 142 university students, of whom 68 were from Eszterházy Károly Catholic University (EKCU) and 74 were from Ludovika University of Public Service (LUPS). Approximately two-thirds of the participants identified as female and one-third as male. The majority (52.1%) were aged 18–25, followed by older age groups ranging from 26 to over 51.

The questionnaire included 24 items across three key flow-related dimensions: Immersion, Balance (challenge–skill match), and AI Support. Each factor was measured on a 0–100% scale using the Artificial Intelligence-Facilitated Flow Questionnaire (AIFLQ). A composite Total Flow Score was calculated as the average of the three sub-scores.

Descriptive statistics were calculated for each factor. For instance, LUPS students achieved an average Immersion score of 83.31% (SD = 12.56), whereas EKCU students achieved an average score of 66.40% (SD = 7.05). These values suggest moderate to high levels of perceived flow during AI-supported learning tasks.

Based on the methodology described above, we conducted an evaluation of students' flow experiences across two universities. The following section presents the aggregated results of this assessment, highlighting key trends and offering insight into the current state of AI-supported learning environments.

The findings indicate a remarkable consistency between the two universities in terms of the AI Flow experience reported by students. Across both institutions, the average Total Flow Score was approximately 68%, as measured by the AIFLQ. This suggests that students experienced a moderate level of flow during AI-supported learning activities, regardless of institutional or contextual differences.

While this score falls below the 80% threshold typically associated with a sustained flow state, it nonetheless reflects a solid foundation upon which further development can build. The results highlight both the current potential and the untapped opportunities in designing AI-enhanced educational environments. These findings emphasize the importance of continued refinement in how AI systems are integrated into learning processes, with the goal of increasing student engagement and reaching higher levels of flow in future implementations (Table 2).

Table 2. Flow score percent in both universities.

Institution	Immersion f. %	Immersion f. s	Balance f. %	Balance f. s	AI f. %	AI f. s	Total Score %	Total Score s
LUPS	83.31	12.56	83.18	10.05	69.42	16.28	78.64	8.16
EKCU	66.40	7.05	73.75	5.03	68.16	5.08	69.44	11.16

Across the two universities, participants achieved an average score of 30.08 (out of 40) on the Immersion factor, 31.46 (out of 40) on the Balance factor, and 27.53 (out of 40) on the AI Support factor. These individual dimensions contributed to a combined Total Flow Score mean of 89.08 (out of 120). These results indicate a generally high level of flow experience among the respondents, suggesting that the AI-supported learning environments were effective in fostering engagement, balance, and meaningful interaction with AI tools (Figures 4–7).

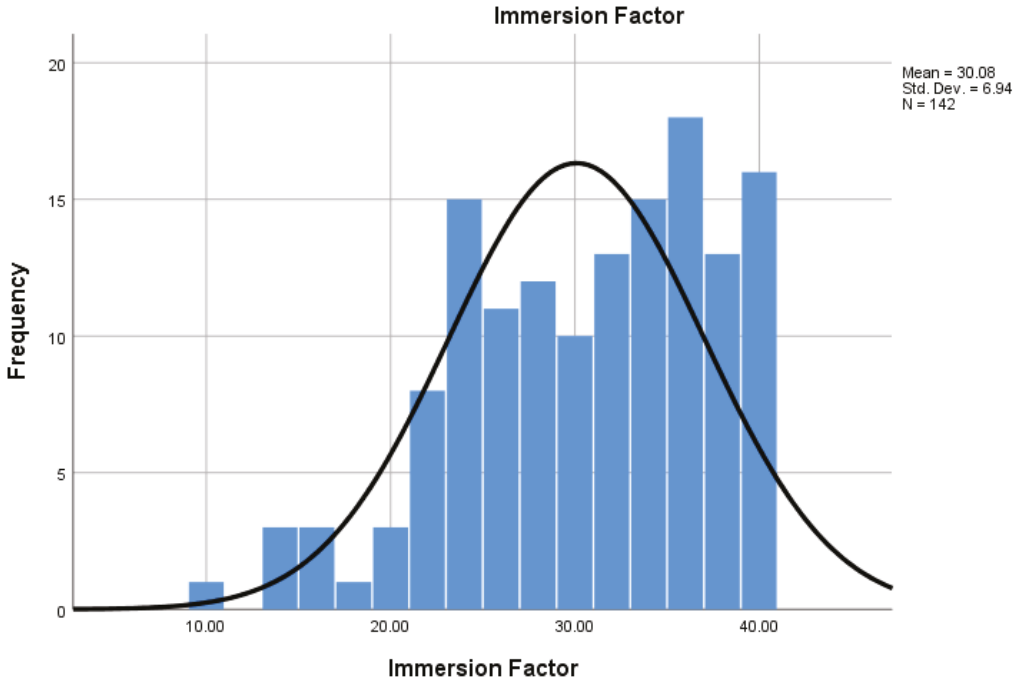


Figure 4. Average score of the Immersion factor.

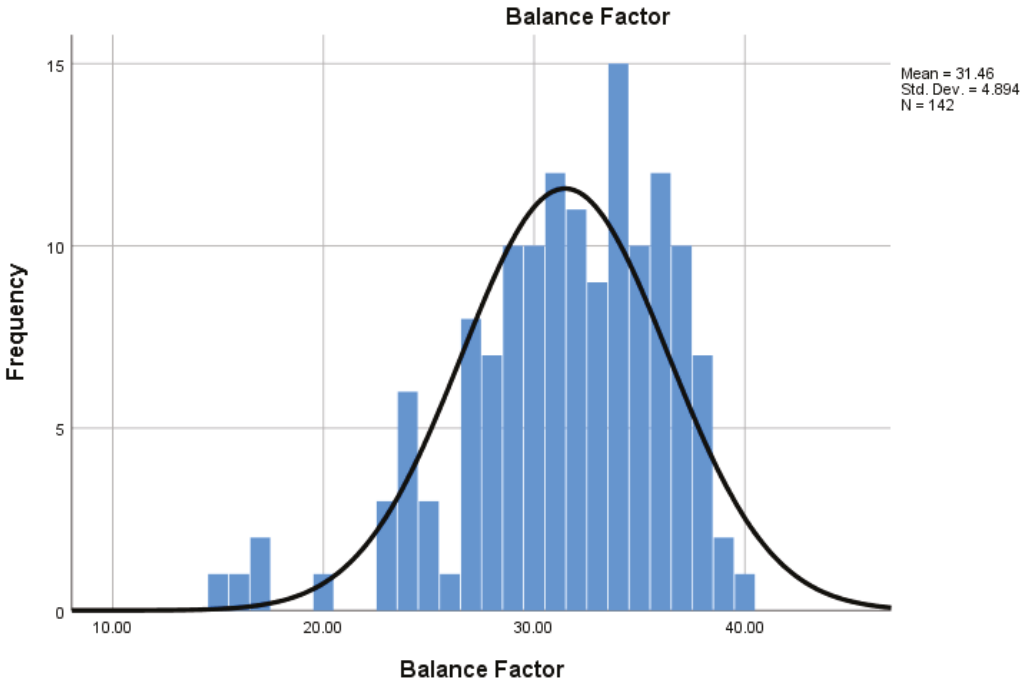


Figure 5. Average score of the Balance factor.

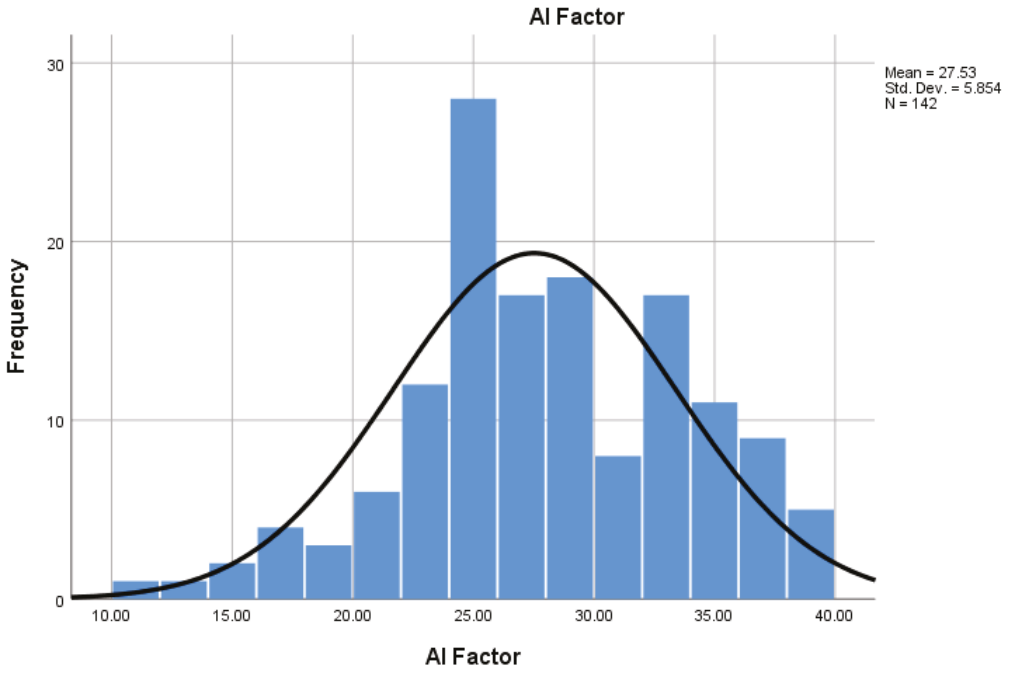


Figure 6. Average score of the AI factor.

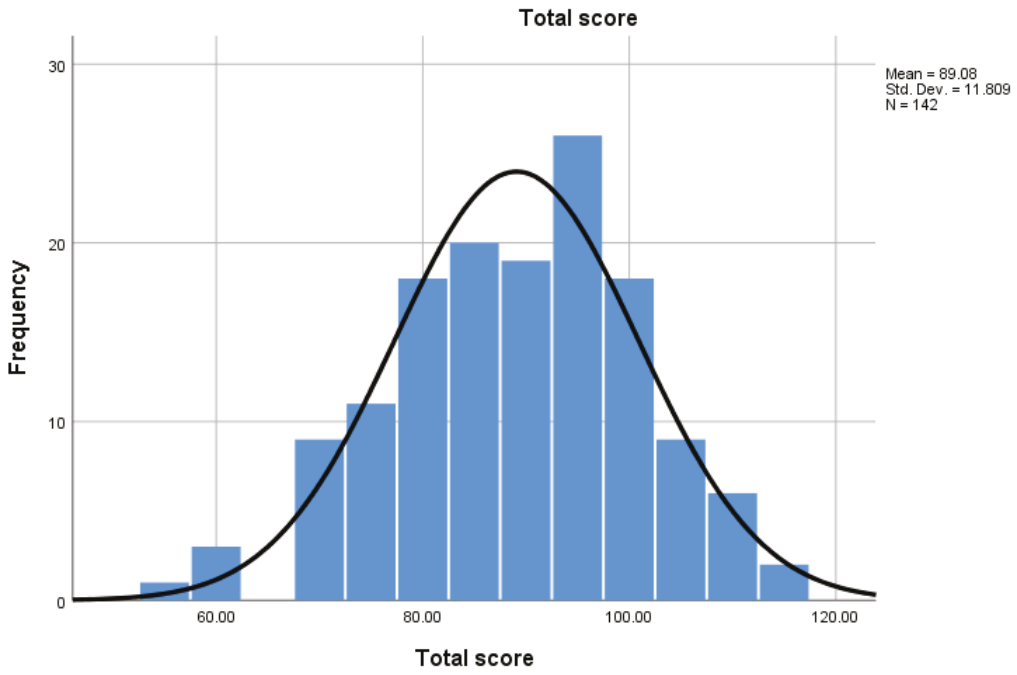


Figure 7. Average total score.

Although the difference in the Immersion factor between male and female participants was not statistically significant in the larger sample (p = 0.059), the Mann–Whitney U test still revealed a trend, with female participants showing higher mean ranks. This tendency, consistent with findings from the earlier, smaller-sample analysis, suggests a potential gender-related difference in how learners experience AI-supported educational tasks. While this result does not meet the conventional threshold for significance, it may indicate that women tend to feel more immersed, warranting further investigation into gender-specific learning experiences with AI tools (Table 3 and Figure 8).

Table 3. Statistical analysis of gender differences in Immersion factor.

Independent-Samples Mann–Whitney U Test Summary (Gender—Immersion Factor)	
Total N	142
Mann–Whitney U	2717.500
Wilcoxon W	7088.500
Test Statistic	2717.500
Standard Error	232.752
Standardized Test Statistic	1.886
Asymptotic Sig. (2-sided test)	0.059

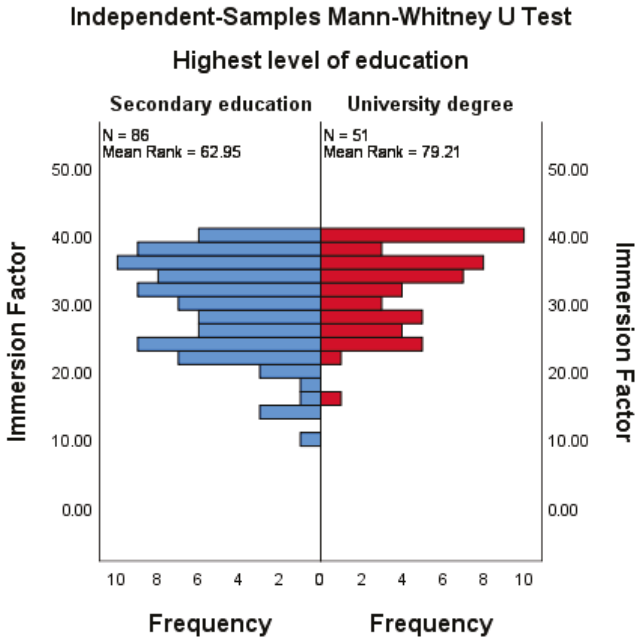


Figure 8. Immersion Scores by Gender.

Further analysis was conducted to examine potential differences in flow experiences across demographic groups. The results revealed a statistically significant difference in the Immersion factor, while the AI Support factor showed a non-significant trend toward group differences. These findings suggest that certain aspects of the flow experience—particularly Immersion—may be influenced by variables such as age. Although the difference in AI Support did not reach statistical significance, the observed pattern still warrants further exploration of how different groups engage with AI-supported learning environments (Table 4).

Table 4. Hypothesis test summary of factors between age groups.

Hypothesis Test Summary				
	Null Hypothesis	**Test**	**Sig.** [a,b]	**Decision**
1	The distribution of Immersion Factor is the same across categories of Age groups.	Independent-Samples Kruskal–Wallis Test	0.002	Reject the null hypothesis.
2	The distribution of Balance Factor is the same across categories of Age groups.	Independent-Samples Kruskal–Wallis Test	0.321	Retain the null hypothesis.
3	The distribution of AI Factor is the same across categories of Age groups.	Independent-Samples Kruskal–Wallis Test	0.198	Retain the null hypothesis.
4	The distribution of Total score is the same across categories of Age groups.	Independent-Samples Kruskal–Wallis Test	0.368	Retain the null hypothesis.

[a]. The significance level is 0.050. [b]. Asymptotic significance is displayed.

Regarding the Immersion factor, the Kruskal–Wallis test revealed a significant difference between the 18–25 and 34–40 age groups, with the 18–25 group showing the lowest mean rank. Although the differences between the 18–25 age group and all other age groups were not statistically significant, it is noteworthy that the 18–25 group consistently had the lowest mean rank across comparisons, suggesting a potential trend (Table 5 and Figure 9).

Table 5. Pairwise comparisons of age groups for the Immersion factor.

Pairwise Comparisons of Age Groups					
Sample 1–Sample 2	**Test Statistic**	**Std. Error**	**Std. Test Statistic**	**Sig.**	**Adj. Sig. [a]**
18–25–26–33	−21.773	9.809	−2.220	0.026	0.264
18–25–41–50	−26.155	10.355	−2.526	0.012	0.115
18–25–34–40	−31.785	10.798	−2.944	0.003	0.032
18–25–51+	−40.373	16.247	−2.485	0.013	0.130
26–33–41–50	−4.382	12.562	−0.349	0.727	1.000
26–33–34–40	−10.012	12.930	−0.774	0.439	1.000
26–33–51+	−18.599	17.736	−1.049	0.294	1.000
41–50–34–40	5.631	13.349	0.422	0.673	1.000
41–50–51+	−14.218	18.043	−0.788	0.431	1.000
34–40–51+	−8.587	18.301	−0.469	0.639	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is 0.050. [a]. Significance values have been adjusted by the Bonferroni correction for multiple tests.

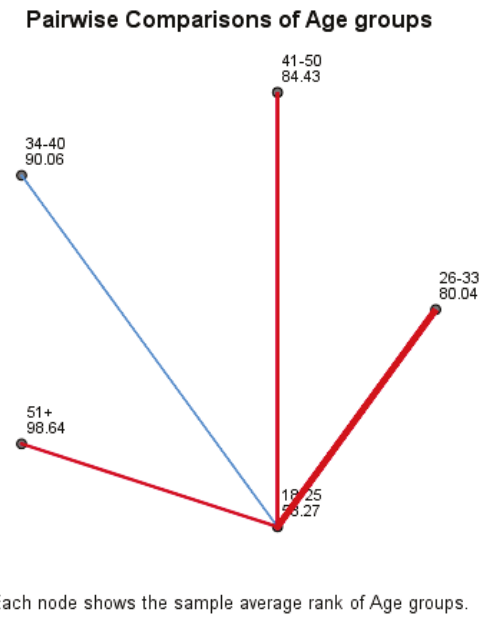


Figure 9. Pairwise comparisons of age groups for the Immersion factor.

While the analysis did not reveal any statistically significant differences between age groups in relation to the AI factor, a noteworthy trend was observed. Specifically, the two youngest age groups (18–25 and 26–33) demonstrated the highest mean ranks among all groups. Although these differences did not reach statistical significance, the consistent elevation in mean ranks among younger participants may indicate a potential age-related pattern in AI-related attitudes or behaviors. This trend, albeit preliminary, suggests that younger individuals might be more receptive to or engaged with aspects measured by the AI factor. Further studies with larger samples may be needed to clarify the nature and significance of this relationship.

A possible explanation for this pattern is that the 18–25 age group, having grown up in a digitally native environment, may be more comfortable and adept at leveraging AI technologies, resulting in a higher appreciation and utilization of AI support. In contrast, the 34–40 age group might be less familiar or less inclined to fully engage with AI-driven tools, which could explain their lower scores. The absence of significant differences among other age groups suggests that this effect may be particularly pronounced only between these two cohorts (Table 6 and Figure 10).

Table 6. Pairwise comparisons of age groups for the AI factor.

Pairwise Comparisons of Age Groups					
Sample 1–Sample 2	**Test Statistic**	**Std. Error**	**Std. Test Statistic**	**Sig.**	**Adj. Sig.** [a]
34–40–41–50	−19.494	13.335	−1.462	0.144	1.000
34–40–51+	−23.159	18.282	−1.267	0.205	1.000
34–40–26–33	25.162	12.916	1.948	0.051	0.514
34–40–18–25	25.870	10.787	2.398	0.016	0.165
41–50–51+	−3.664	18.024	−0.203	0.839	1.000
41–50–26–33	5.667	12.549	0.452	0.652	1.000
41–50–18–25	6.376	10.344	0.616	0.538	1.000
51+–26–33	2.003	17.717	0.113	0.910	1.000
51+–18–25	2.711	16.230	0.167	0.867	1.000
26–33–18–25	0.708	9.798	0.072	0.942	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is 0.050. [a]. Significance values have been adjusted by the Bonferroni correction for multiple tests.

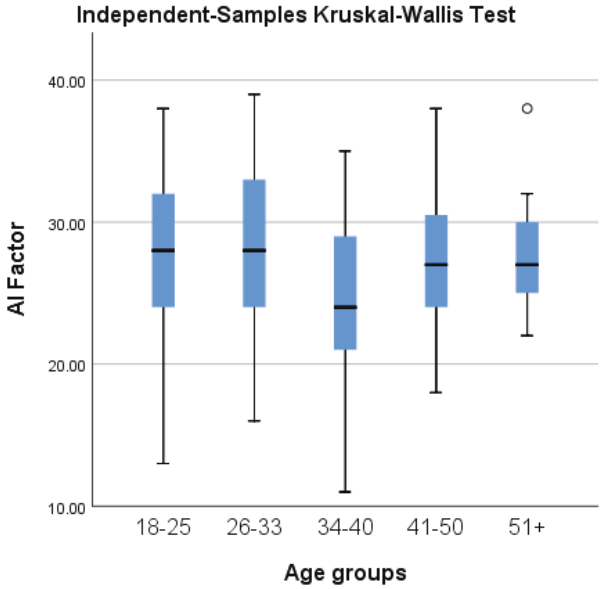


Figure 10. Kruskal-Wallis Test of age groups for AI factor.

However, statistical analysis revealed no significant correlations between participants' self-reported academic orientation (Humanities and Social Sciences, Public Administration and Legal Studies, STEM, and Environmental Sciences) and any of the measured flow factors —Immersion, Balance, AI—or the Total Flow score. This suggests that the experience of flow in AI-supported learning environments is consistent across different personality or disciplinary orientations.

In addition to the observed age-related differences, no significant correlations were found between the flow factors (Immersion, Balance, AI, Total Flow score) and participants' level of educational attainment. One possible explanation for this lack of association is

the relatively homogeneous age distribution within the sample, which may have limited variability in educational background and thus reduced the potential to detect meaningful differences related to educational levels.

Based on the above, and going back to our four hypotheses formulated earlier, we can make the following observations:

H1. *The use of AI-based learning support is positively associated with the experience of flow among university students. The average AI flow factor score across both universities was approximately 68–69%, which, while below the 80% threshold used to define a sustained flow experience, still indicates a moderate presence of flow. This suggests that students did experience a certain degree of engagement and absorption in AI-supported learning environments. Although this result does not fully confirm the hypothesis, it offers partial support, indicating that AI integration in learning has the potential to facilitate flow, but further development or contextual adaptation may be necessary to increase its impact.*

H2. *Female students report significantly higher scores on the immersion factor of flow compared to male students. This hypothesis was supported by the data. Female participants achieved significantly higher mean ranks in the immersion factor, suggesting they experienced a deeper sense of absorption and engagement in the learning tasks. This finding is consistent with previous research indicating potential gender differences in emotional or cognitive engagement during learning, and it highlights the importance of considering gender-responsive approaches when designing AI-supported educational tools.*

H3. *Students with higher educational qualifications demonstrate significantly higher overall flow scores than those with lower qualifications. This hypothesis was not fully supported. While no significant relationship was found between educational attainment and overall flow scores, a statistically significant association did emerge for the immersion factor. Based on the distribution of responses, the analysis focused on students with either secondary school or university degrees. Interestingly, those with a university degree exhibited significantly higher mean ranks on the immersion factor compared to those with a secondary school qualification. This finding suggests that, in this sample, students with higher formal educational attainment may have experienced greater immersion. One possible explanation could involve the influence of academic experience on the ability to engage deeply with learning tasks, although further research is needed to explore this effect in more depth (Table 7 and Figure 8).*

Table 7. Statistical analysis of educational level differences across all factors and total score.

Hypothesis Test Summary				
	Null Hypothesis	**Test**	**Sig.** [a,b]	**Decision**
1	The distribution of Immersion Factor is the same across categories of Highest level of education.	Independent-Samples Mann–Whitney U Test	0.020	Reject the null hypothesis.
2	The distribution of Balance Factor is the same across categories of Highest level of education.	Independent-Samples Mann–Whitney U Test	0.718	Retain the null hypothesis.
3	The distribution of AI Factor is the same across categories of Highest level of education.	Independent-Samples Mann–Whitney U Test	0.257	Retain the null hypothesis.
4	The distribution of Total score is the same across categories of Highest level of education.	Independent-Samples Mann–Whitney U Test	0.464	Retain the null hypothesis.

[a]. The significance level is 0.050. [b]. Asymptotic significance is displayed.

H4. *Personality type has no significant effect on the experience of flow among university students. This hypothesis was confirmed by the findings. Participants' self-reported academic orientation or personality profile (e.g., Humanities and Social Sciences, Public Administration and Legal*

Studies, STEM, and Environmental Sciences) showed no significant correlation with any of the flow dimensions. This suggests that the capacity to experience flow in AI-supported learning environments is not inherently tied to personality traits or disciplinary background, reinforcing the idea that well-designed technological tools can be broadly effective across diverse learner profiles.

Although hypothesis H4—that personality type has no significant effect on the experience of flow—did not show a statistically significant correlation, it is essential to examine this result in order to gain a full understanding of the phenomenon. The non-significant result does not automatically exclude the role of personality or study orientation in the experience of flow; rather, it reflects the fact that, within the framework of the present study and given the sample and methodological approach examined, no clear evidence of such an effect was found. However, this finding has important implications: it suggests that the experience of flow in AI-supported learning environments depends on the quality of the learning environment and the effectiveness of the technology used rather than on personal characteristics or field of study. This conclusion is significant because it implies that well-designed, adaptive, and interactive technological tools have the potential to facilitate the experience of flow for a diverse range of learners, regardless of their personality traits or disciplinary background. Thus, the results support the inclusive nature of technology in education and contribute to the development of more effective personalized learning environments. Therefore, this non-significant result is valuable evidence that AI-based learning tools can be effective beyond traditional personality or domain categories, helping to improve the learning experience and broaden the reach of flow.

4. Discussion

The findings of this study provide important insights into how university students experience flow in AI-supported learning environments. The partial confirmation of our first hypothesis suggests that while the integration of artificial intelligence tools does support the emergence of flow, it does not yet reach the level of sustained engagement typically associated with high flow states. An average AI flow score of approximately 68–69% indicates a moderate level of engagement and a promising baseline for further development.

Notably, gender emerged as a significant factor in students' immersion levels. Female participants reported significantly higher immersion scores than their male counterparts, aligning with prior research suggesting that women may engage more deeply in structured, cognitively demanding learning contexts. These findings emphasize the importance of gender-sensitive instructional design in AI-enhanced education.

In contrast, self-identified personality type showed no significant relationship with flow experiences. However, a significant association was found between educational level and the immersion component of flow, with students holding a university degree reporting higher levels of immersion than those with a secondary school qualification. This finding suggests that while overall flow may not be strongly tied to academic background, certain aspects of the experience—such as immersion—could be influenced by educational level. It is possible that students with higher educational qualifications are better equipped to engage deeply with learning tasks, perhaps due to greater academic experience or more developed learning strategies. Nonetheless, the broader pattern still indicates that flow is shaped more by the immediate learning context and instructional design than by stable individual characteristics. The relative age homogeneity of the sample may also have contributed to the lack of significant associations across other dimensions.

Interestingly, age-related differences were observed in both the immersion and AI support dimensions. The youngest age group (18–25) scored significantly lower on immersion, possibly due to digital saturation, multitasking habits, or a reduced novelty effect

of AI tools. Conversely, this same group reported significantly higher AI support scores compared to older participants, which may reflect greater familiarity with and trust in AI technologies. These age-specific dynamics point to the complex and potentially ambivalent relationship between digital nativity and deep engagement in learning.

5. Conclusions

This study contributes to the growing body of literature on flow experiences in AI-supported learning, yet several limitations must be acknowledged. First, the sample was drawn from two universities and included a relatively age-homogeneous population, which may have restricted the variability needed to detect certain effects, particularly regarding educational attainment. Additionally, flow was assessed using self-report measures, which, while validated, are inherently subjective and may be influenced by participants' current mood or interpretation of the items.

Another limitation is the reliance on participants' self-classified personality or academic orientation rather than standardized personality assessments, which may limit the interpretability of findings related to learner traits.

Despite these constraints, this study reveals valuable patterns. AI-supported learning environments show potential in fostering moderate levels of flow, particularly among digitally fluent learners. However, a more nuanced design approach may be needed to increase immersion and engagement, especially for younger students who may not find standard AI features sufficiently novel or stimulating. Gender differences in immersion also highlight the need for differentiated learning strategies.

In conclusion, while not all hypotheses were fully confirmed, the findings underscore the importance of tailoring AI-driven educational experiences to diverse learner profiles, taking into account variables such as gender and age. Future research should build on these insights by incorporating longitudinal designs, a broader range of demographic backgrounds, and objective measures of engagement to further refine our understanding of flow in technologically enhanced learning environments.

From a practical perspective, the findings of this study provide educators, curriculum developers and edtech designers with actionable insights. By identifying the influence of factors such as gender and educational attainment on the experience of flow in AI-supported learning environments, instructors can better tailor their interventions to maximize engagement. For example, targeted scaffolding or differentiated instruction could be particularly beneficial for groups of students who are less likely to achieve immersion. Furthermore, the use of validated instruments such as the AIFLQ offers a scalable means of assessing and refining AI-based educational strategies on an ongoing basis.

To further refine our understanding of flow in technologically enhanced learning environments, future research should build on these insights by incorporating longitudinal designs, a broader range of demographic backgrounds, and objective measures of engagement.

Author Contributions: D.L.D. and Z.S.; methodology, D.L.D. and S.C.; software, D.L.D. and S.C.; validation, T.L.M. and R.R.; formal analysis, D.L.D. and S.C.; investigation, D.L.D., Z.S. and D.L.D.; data curation, D.L.D.; writing, D.L.D. and G.V.; visualization, S.C.; supervision, Z.S.; project administration, G.V.; funding acquisition, D.L.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Csíkszentmihályi, M. *Flow: The Psychology of Optimal Experience*; Harper & Row: New York, NY, USA, 1990.
2. Shernoff, D.J.; Csíkszentmihályi, M.; Schneider, B.; Shernoff, E.S. Student engagement in high school classrooms from the perspective of flow theory. *Sch. Psychol. Q.* **2003**, *18*, 158–176. [CrossRef]
3. Pintrich, P.R. A conceptual framework for assessing motivation and self-regulated learning in college students. *Educ. Psychol. Rev.* **2004**, *16*, 385–407. Available online: https://link.springer.com/article/10.1007/s10648-004-0006-x (accessed on 8 July 2025). [CrossRef]
4. Keller, J.M. First principles of motivation to learn and e3-learning. *Distance. Educ.* **2008**, *29*, 175–185. [CrossRef]
5. Martens, R.; Gulikers, J.; Bastiaens, T. The impact of intrinsic motivation on e-learning in authentic computer tasks. *J. Comput. Assist. Learn.* **2004**, *20*, 368–376. [CrossRef]
6. Farkas, I.; Kővári, A.; Rajcsányi-Molnár, M. The emergence of artificial intelligence in education and its impact on individual literacy in higher education. In Proceedings of the 2024 IEEE 7th International Conference and Workshop Óbuda on Electrical and Power Engineering (CANDO-EPE), Piscataway, NJ, USA, 17–18 October 2024; IEEE: Piscataway, NJ, USA, 2024; Volume 347, pp. 83–88.
7. Luckin, R.; Holmes, W. *Intelligence Unleashed: An Argument for AI in Education*; UCL Knowledge Lab: London, UK, 2016.
8. Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. Systematic review of 11research on artificial intelligence applications in higher education—Where are the educators? *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 39. [CrossRef]
9. Holmes, W.; Bialik, M.; Fadel, C. *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Center for Curriculum Redesign*; Center for Curriculum Redesign: Boston, MA, USA, 2019; ISBN 978-1794293700.
10. Dominek, D.L. Development and Validation of the AI and Flow Learning Questionnaire (AIFLQ). *J. Appl. Tech. Educ. Sci.* **2025**, *15*, 1–20. Available online: https://jates.org/index.php/jatespath/article/view/399/188 (accessed on 8 July 2025).
11. Khristin, F.; Topping, K.J.; Barron, G. Using mobile technologies for mathematics: Effects on student attitudes and achievement. *Educ. Technol. Res. Dev.* **2018**, *66*, 1119–1139. [CrossRef]
12. Shahzad, M.F.; Xu, S.; Zahid, H. Exploring the impact of generative AI-based technologies on learning performance through self-efficacy, fairness & ethics, creativity, and trust in higher education. *Educ. Inf. Technol.* **2025**, *30*, 3691–3716. [CrossRef]
13. Huang, F.; Peng, D.; Teo, T. AI Affordances and EFL Learners' Speaking Engagement: The Moderating Roles of Gender and Learner Type. *Eur. J. Educ.* **2025**, *60*, e70041. [CrossRef]
14. Hwang, G.J.; Xie, H.; Wah, B.W.; Gašević, D. Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Comput. Educ. Artif. Intell.* **2020**, *1*, 100001. [CrossRef]
15. Shamir-Inbal, T.; Blau, I. Characteristics of pedagogical change in integrating digital collaborative learning and their sustainability in a school culture: E-CSAMR framework. *J. Comput. Assist. Learn.* **2021**, *37*, 825–838. Available online: https://onlinelibrary.wiley.com/doi/10.1111/jcal.12526?msockid=34debf28c52a69151156aaf9c4f368c1 (accessed on 8 July 2025).
16. Szűts, Z.; Molnár, G.; Racsko, R.; Vaughan, G.; Molnár, L. Pedagogical Implications and Methodological Possibilities of Digital Transformation in Digital Education after the COVID-19 Epidemic. *Computers* **2023**, *12*, 1–19. [CrossRef]
17. D'Mello, S.K.; Graesser, A. AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Trans. Interact. Intell. Syst. (TiiS)* **2012**, *2*, 1–39. Available online: https://dl.acm.org/doi/10.1145/2395123.2395128 (accessed on 8 July 2025). [CrossRef]
18. Duckworth, A.L.; Peterson, C.; Matthews, M.D.; Kelly, D.R. Grit: Perseverance and passion for long-term goals. *J. Personal. Soc. Psychol.* **2007**, *92*, 1087–1101. [CrossRef] [PubMed]
19. Zimmerman, B.J. Becoming a self-regulated learner: An overview. *Theory. Into. Pract.* **2002**, *41*, 64–70. Available online: https://www.tandfonline.com/doi/abs/10.1207/s15430421tip4102_2 (accessed on 8 July 2025). [CrossRef]
20. Arató, B.; Balázs, G. The communicative-linguistic modes of artificial intelligence with a focus on justice. *Magy. Nyelvőr.* **2024**, *148*, 601–617. [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.

computers

MDPI

Article

Bridging Tradition and Innovation: Transformative Educational Practices in Museums with AI and VR

Michele Domenico Todino [1,*], Eliza Pitri [2], Argyro Fella [2,*], Antonia Michaelidou [2], Lucia Campitiello [1], Francesca Placanica [3], Stefano Di Tore [1] and Maurizio Sibilio [1]

[1] Department of Humanities, Philosophy and Education/DISUFF, University of Salerno, 84084 Fisciano, Italy; lcampitiello@unisa.it (L.C.); sditore@unisa.it (S.D.T.); msibilio@unisa.it (M.S.)
[2] School of Education, University of Nicosia, Nicosia 1700, Cyprus; pitri.e@unic.ac.cy (E.P.); michaelidou.a@unic.ac.cy (A.M.)
[3] Department of Law, Economics and Human Sciences, University of Reggio Calabria, 89124 Reggio Calabria, Italy; f.placanica@unimc.it or francesca.placanica@unirc.it
* Correspondence: mtodino@unisa.it (M.D.T.); fella.a@unic.ac.cy (A.F.)

Abstract: This paper explores the intersection of folk art, museums, and education in the 20th century, with a focus on the concept of art as experience, emphasizing the role of museums as active, inclusive learning spaces. A collaboration between the University of Salerno and the University of Nicosia has developed virtual museum environments using virtual reality (VR) to enhance engagement with cultural heritage. These projects aim to make museums more accessible and interactive, with future potential in integrating artificial intelligence NPC and VR strategies for personalized visitor experiences of the Nicosia Folk Art Museum.

Keywords: museum education; folk art; virtual museums; AR/VR; IA; inclusion

1. Introduction

Folk culture consists of essential elements that influence the structure of our society, with various traditional aspects continuing to persist and affect how people view the world today [1]. Folk culture serves as a dynamic testament to the evolving cultural identity of each community, shaped by historical, environmental, and social influences over time. Traditions, while transmitted across generations, are not static; rather, they are continually reinterpreted to reflect the specific needs and circumstances of each era. The capacity of folk culture to adapt ensures its relevance across varying temporal and cultural contexts. For any traditional form of expression to maintain its significance and resonance within contemporary society, it must undergo a transformation that aligns with shifting cultural narratives and societal expectations. Such evolution is crucial to its survival; without it, traditional expressions risk becoming obsolete, disconnected from the values and lived experiences of modern audiences [2].

The intersection of tradition, transition, and transformation plays a pivotal role in contemporary educational practices [3–5], particularly in the context of folk art and museums. In order to deepen students' understanding of the diversity, cultural significance, historical evolution, and social functions of folk art forms, educators seek to foster an authentic engagement with and appreciation for these artistic traditions. Within the framework of critical museology, museums are increasingly re-evaluating their social function, positioning themselves as spaces for knowledge exchange and active participation in forging connections between tradition and contemporary experience [6]. This shift reflects a

broader transition in the museum's role—from a repository of static heritage to a dynamic forum for cultural dialog. As a result, museum education is in constant need of a significant transformation, emerging as an autonomous and reflective practice that challenges, expands, and redefines both exhibitions and institutional structures to promote meaningful learning experiences for a diverse audience of visitors [7]. This paper aims to explore the development of virtual museum environments for the Nicosia Folk Art Museum through the use of virtual reality (VR) technologies. It discusses how these educational initiatives can enhance engagement with cultural heritage by making folk art more accessible, interactive, and meaningful for contemporary audiences. Framed within the context of museology, this paper investigates how VR can support the museum's evolving role as a dynamic space for cultural dialog, education, and personalized visitor experiences—bridging the intersection of tradition, transition, and transformation in museum education.

2. The Example of an Educational Project Developed with Non-Digital Technologies

Before delving into the technological dimensions of these educational initiatives, it is essential to describe the background and the context that inspired the design of the software. Museums are vital spaces for engagement, promoting accessibility, inclusivity, and transformation [8,9]. A key example is the collaboration between the UNESCO Chair and the University of Nicosia's Department of Education at the Nicosia Folk Art Museum. Established in 1937 in the Old Archbishop's Palace, the Nicosia Folk Art Museum (https://www.nicosia.org.cy/en-GB/discover/museums/cyprus-folk-art-Museum/ accessed on 22 June 2025) preserves Cypriot folk art, reflecting the island's cultural identity through diverse forms of folk art, such as weaving, pottery, woodcarving, metalwork, basketry, leatherwork embroidery, lace, etc. The objective of the educational initiative, which has already been carried out, at the Nicosia Folk Art Museum was to engage preschool children through hands-on experiences led by student-teachers. An event at the Nicosia Folk Art Museum commenced with a short theatrical performance devised and enacted by the student-teachers, based on a reimagined Cypriot folk myth titled The Naughty Gnome of the Well. This adaptation playfully transformed the traditional tale of haunting gnomes into a child-friendly narrative in which the mischievous gnome, after stealing pretzels from a baker, falls into a well. Despite the gnome's antics, the fictional community expressed a sense of loss in its absence, prompting the community leader—within the narrative—to challenge the characters and children to complete four tasks in order to bring the gnome back. This storytelling performance provided a narrative framework for guiding the children through four interactive activity stations, each designed around the pedagogical concept of patterns (visual, auditory, and kinesthetic). Each activity was intentionally designed to foster experiential learning and promote core developmental competencies. Throughout these engagements, the children collected handkerchiefs representing abstract values—knowledge, cooperation, humor, and patience—associated with each activity. At the conclusion of the event, these symbolic artifacts were joined together during a celebratory group dance marking the return of the gnome, thereby reinforcing the event's overarching themes. Guided by new museology principles, which emphasize context and social relevance, the project encouraged an appreciation of folk art while transforming the museum into an active learning space.

2.1. Educational Project Outcomes and Benefits

Artistic creation entails addressing both technical and conceptual challenges, positioning the creative process itself as a form of problem solving. Technical challenges involve mastering the use of materials and media—such as techniques for painting or

constructing—while conceptual challenges relate to the development and communication of underlying meanings and ideas, such as determining what subject matter to represent. The artist navigates choices concerning media, design principles, the thematic content, and stylistic approaches, all within a framework guided by a central conceptual focus. Rather than constraining the process, practical and conceptual boundaries serve to sharpen the intent and enrich the creative outcome, supporting a focused and purposeful artistic endeavor [10]. Conceptual problem solving in art extends beyond the initial articulation of a big idea, requiring an ongoing inquiry into its multiple dimensions. This may involve strategies such as omitting information, subverting conventional perspectives, or juxtaposing disparate elements—objects, figures, or contexts—in novel ways. Transformation emerges as a central strategy, wherein artists deliberately alter standard representations using visual means to convey deeper or unexpected meanings. In resolving such conceptual issues, artists engage in cognitive practices such as risk-taking, imaginative play, exploration, and the use of humor. These strategies enable them to push beyond traditional limits, generating innovation and enhancing the expressive power of their work [11].

Similarly, educators encounter complex challenges in the instructional design process, particularly after a central theme or big idea has been established. Decisions must be made regarding pedagogical methods, activity structures, and the selection of media and materials in a manner that aligns with the learning objectives, the educational environment, and the developmental needs of young learners. Humor is identified as a valuable tool in facilitating the transformation of educational practices. However, it must be employed with care; while humor can provoke laughter, the two are not synonymous, and not all forms of humor are suitable for educational settings. Humor that is manipulative, demeaning, or discriminatory is inappropriate, especially in early childhood contexts. For young children, laughter is frequently elicited by the unexpected or incongruous—elements that challenge established norms and expectations, offering playful yet meaningful opportunities for engagement and reflection [12].

Humor emerged as a significant pedagogical element contributing to the success of the event at the Cyprus Folk Art Museum, offering notable benefits for both the student-teachers and the participating children. More broadly, humor facilitates positive interpersonal interactions and strengthens social bonds. In educational contexts, the strategic use of humor has been shown to enhance the overall classroom climate, promote effective communication between educators and learners, and serve as an informal classroom management technique that supports constructive teacher–student relationships. When appropriately aligned with the instructional content, humor can diminish critical judgment, reduce stress, and sustain student attention and engagement by fostering a more enjoyable learning experience. Additionally, humor has been linked to the development of critical thinking, creativity, imagination, and a sense of community among learners. Importantly, educators are not required to possess professional comedic skills to effectively integrate humor into their teaching practice. Simple yet intentional strategies—such as opening lessons with light-hearted jokes, embedding humorous prompts in assessments, or contextualizing content through playful scenarios like The Naughty Gnome of the Well—can stimulate student interest. Furthermore, incorporating unconventional methods, such as using song lyrics to illustrate mathematical concepts or introducing unexpected, respectful surprises, can serve to captivate attention and enhance the learning experience in meaningful ways.

A central component of the educational event at the Cyprus Folk Art Museum was the strategic incorporation of storytelling. Employed as a pedagogical framework, storytelling effectively structured the sequence of activities and guided children toward meaningful and purposeful engagement. When integrated with contemporary digital technologies, storytelling holds considerable potential to stimulate children's imagination and deepen

their engagement with the cultural content. For instance, traditional motifs—such as patterns from a woven bedcover—can be digitally animated and transformed into imaginative, unexpected forms, thus enriching the interpretive possibilities of heritage artifacts (see Figure 1).

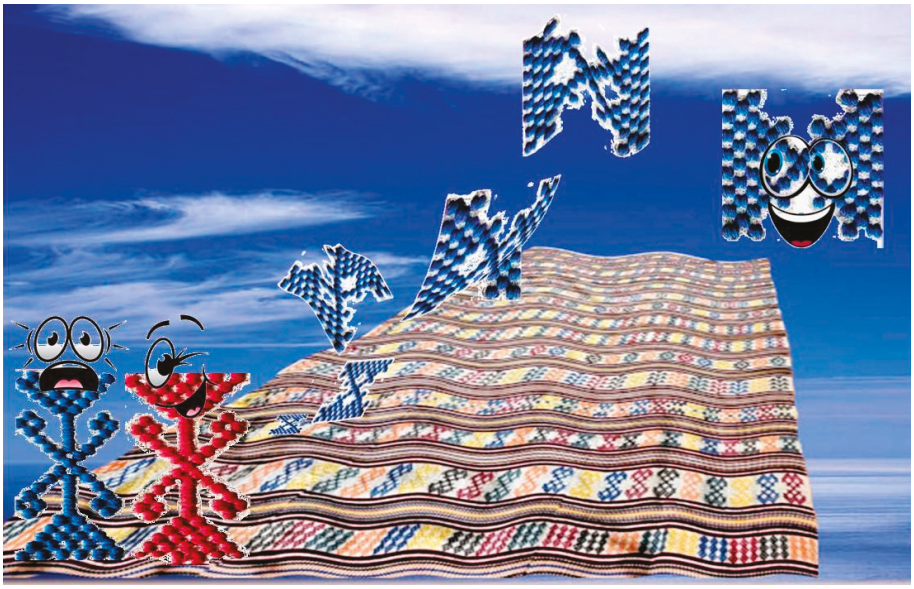


Figure 1. The figure illustrates the design derived from previous prototypes of the educational game related to the traditional Cypriot tablecloth in augmented reality.

2.2. Expanding the Educational Project Through an AI-Based Tutor to Make Folk Art Relevant and Increase Accessibility

The aforementioned educational project was performed with non-digital technologies, now we would say unplugged, which was fundamental because it allowed a fruitful hybridization with new technologies seen as a playful–didactic environment.

However, emerging technologies, including virtual and augmented reality, offer innovative means for reinterpreting tradition in ways that resonate with younger audiences. These tools facilitate immersive learning experiences that can enhance children's understanding of cultural heritage and historical processes. Educators might, for example, employ digital storytelling to trace the journey of a single thread—from its movement in the wind, through the stages of spinning and weaving, to its final transformation into a textile object exhibited in the museum. Such an approach enables children to embark on a conceptual journey that links the past with the present and projects into the future, thus fostering historical consciousness and imaginative thinking.

Emerging technologies have opened up significant opportunities for museums across various dimensions, providing new and alternative methods for engaging with visitors. From reconstructing historical scenes to enhancing interpretation, education, and immersion both on-site and remotely [13,14], emerging technologies foster greater accessibility and interaction. Many museums now use these tools to democratize collections and support multiple interpretations, with "accessibility, inclusion, and democratization" recognized as major benefits [15]. As Chiscano and Jiménez-Zarco [16] argue, digital tools that remove access barriers and offer adaptive formats are essential for promoting equal opportunities for all learners, including individuals with disabilities. Therefore, the development of accessible digital materials related to folk and popular culture is not only pedagogically valuable but also socially imperative. By enabling inclusive engagement with traditional arts and narratives, such resources contribute to a more equitable and culturally responsive educational landscape.

Avatar-based digital storytelling plays a significant role in advancing diversity, equity, accessibility, and inclusion in both cultural and educational contexts. It supports these goals in several key ways. First, it offers various formats for information delivery—such as text, audio, and video—catering to diverse learning styles and cognitive needs. Second, interactive storytelling enhances engagement by providing immersive, personalized experiences that users can navigate at their own pace. Third, it promotes a safe and inclusive environment through universal design principles, enabling individuals to manage their interaction with the content. Furthermore, accessibility features like closed captioning and audio descriptions ensure that individuals with a broad spectrum of abilities can fully engage. By allowing users to tailor their experiences to meet their specific needs and interests, avatar-based digital storytelling emerges as a potent tool for cultivating inclusive and engaging educational environments [17].

3. Educational Technologies and Museum Transformation

Since the 1960s, museums have gradually adopted digital technologies, from early computer use to the widespread digitization driven by the Internet in the 2000s. The UNESCO [10] highlights how museums adapted during the pandemic to maintain public engagement, with an emphasis on digital inclusion and accessibility. While their core functions—acquisition, preservation, research, exhibition, and communication—remain unchanged, digitization has introduced new modes of production, distribution, and consumption [8,11]. These shifts, in public engagement and participatory services related to museums' digital content need further exploration.

The design of a museum edugame has to take into account how the museum design, museology, the subject of design, and the teaching–learning process in the present has to become an appropriate human–machine interaction. This paragraph aims to provide a brief introductory overview of how art, museums, and education intersected throughout the twentieth century. To illustrate this, we will consider, among many relevant authors, a fundamental figure: John Dewey. The concept of *art as experience* constitutes one of the most significant contributions of John Dewey [18]. He goes beyond developing a philosophy of art as a specialized discipline. Instead, a new paradigm for experience in general emerges, identified in the esthetic as an order of the refinement and fulfillment of human sensory and emotional impulses.

This perspective entails a keen awareness to critically grasp and evaluate the dense relationships between art and the esthetic, both within experiential dynamics and the social realities that "nourish" them. Dewey thus illuminates the crises of contemporary culture, outlining the framework of an anthropology enriched by the fruits of early twentieth-century artistic revolutions. These characteristics have established *Art as Experience* as a classic in reflections on esthetic experience, influential to significant artists from Josef Albers to Mark Rothko, and today is still central to various crucial domains of philosophical debate [19].

We may assert that the museum occupies an ideal locus for direct engagement, wherein it transcends into a realm of vitality, accessibility, inclusivity, and transformation. A theoretical underpinning, starting with Dewey's progressive and dynamic educational philosophy, serves as a robust paradigmatic foundation. While luminaries like Hein [19] or Bruner [20] could be invoked, it behooves us, in this instance, to narrow our focus and accentuate the convergent essence at hand. The role of preserving cultural assets, both material and immaterial, is accompanied by a significant social function as guardians of a people's narrative, drawing on the potential of ethnology and anthropology, as noted by Clifford in *Narrare humanum est* [21]. Museums thus become spaces fostering a sense of belonging and offering unique opportunities to share experiences and learn together

through direct engagement with artworks and cultural artifacts, in a continuum of different learning environments where constructivism can make a difference [22].

Having established this initial premise, it is now necessary to introduce a second one. In the laboratory of the *Teaching Learning Center for Education and Inclusive Technologies: Elisa Frauenfelder*, affiliated with the Department of Human, Philosophical, and Educational Sciences at the University of Salerno, the development of digital museum environments has been underway since 2016. The aim is to allow students, guided by their teachers, to take on the role of museum curators, applying fundamental principles of exhibiting collections of artifacts and artworks according to new museology criteria.

In this context, a collaborative initiative has been established between the University of Salerno, Italy, and the University of Nicosia, Cyprus, specifically within their Departments of Education and Schools of Education. This collaboration has already initiated an Erasmus+ Mobility Agreement for staff teaching mobility between the two institutions. What exactly is a digital museum environment, and how should it be developed, according to which principles? It is important to note that such environments adhere to the principles of new museology, which emphasizes the contextualization of artworks and the social role of the institution [23] (p. 84), and also, in line with the best practices of real museums [24], for example, it should aim "to elicit the emergence of particular visitor role identities" [25] (p. 10). In other words, it could be argued, therefore, that the museum becomes a place not of passivity but of active learning that is customized around the visitor and his or her specific interests and learning styles. In accordance with the long-standing hypotheses proposed by Guerra, Caprara, and Scala [26], which have been established for nearly two decades, the digital museum can offer experiences and stimulate activities that the physical museum cannot propose due to cost constraints or practical impossibilities.

For instance, physical objects cannot be touched, whereas digital simulations or guided activities offer alternative forms of engagement. Educational software and guided exploration tools allow for the expansion of the cultural and perceptual experience of the visit. A virtual museum can also facilitate distance learning when a virtual museum is available online (see Figures 2–5). Among the various ongoing projects, one in particular deserves a mention. This paper focuses on the implementation of a digital educational resource centered on the traditional Cypriot blanket, as mentioned above.

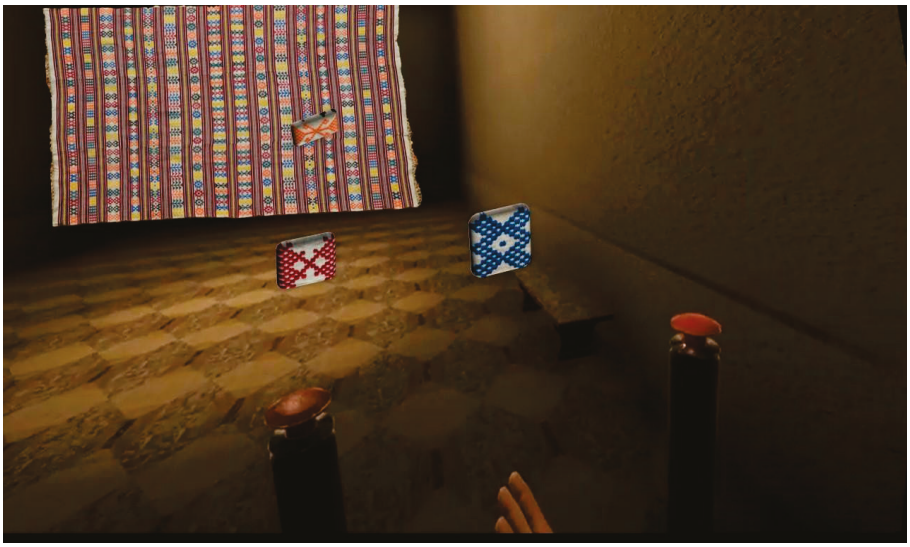


Figure 2. The figure illustrates the first design derived from previous prototypes of the educational game related to the traditional Cypriot blanket in virtual reality.

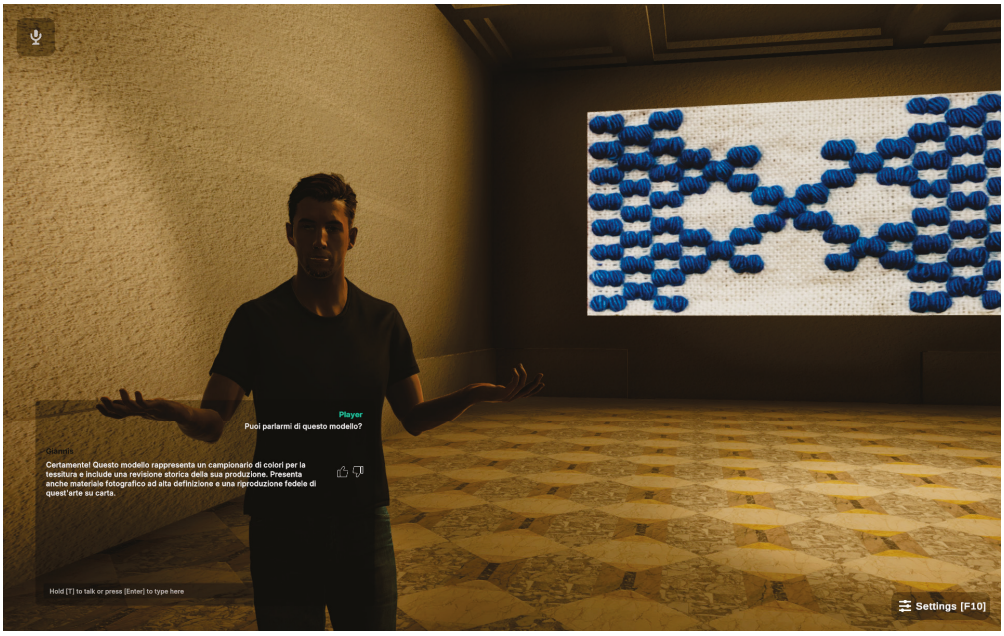


Figure 3. The NPC answers in Italian to a visitor's question in the virtual museum.

Figure 4. The NPC answers in English to a visitor's question in the virtual museum.

Recent studies have highlighted how virtual and immersive technologies are reshaping the relationship between museums, collections, and audiences. A bibliometric analysis by Wang et al. [27] identified five thematic clusters in the application of VR and AR to museum exhibitions, including immersive storytelling, digital heritage, and visitor-centered design. These contributions emphasize the need for meaningful, not merely technological, innovation—pointing to a gap in the research integrating AI, personalization, and pedagogical goals. Similarly, Shehade and Stylianou-Lambert [28] explore the perceptions of museum professionals, who recognize the educational and affective value of VR but raise concerns about user isolation, costs, and the lack of cohesive strategies.

The design proposed in this paper addresses those gaps by integrating a conversational AI agent and VR experience into a pedagogically grounded museum pathway. By combining immersive storytelling, personalized interaction, and remote accessibility, the

Cyprus Folk Art Museum initiative aligns with international best practices and responds to emerging challenges in the field.

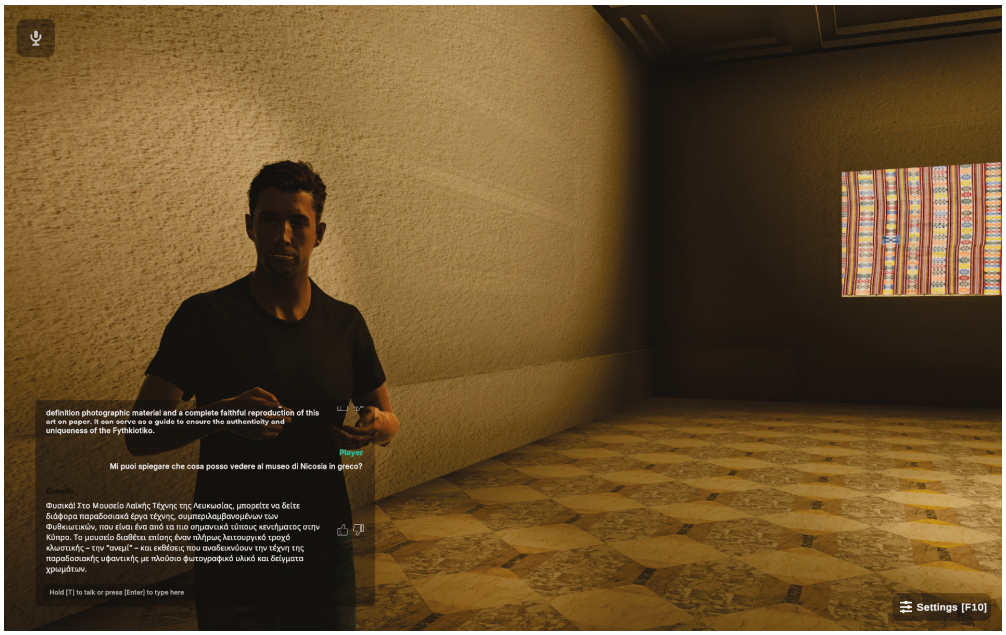


Figure 5. The NPC answers in Greek to a visitor's question in the virtual museum.

4. Case Study 1: Integrating AI for Accessibility and Engagement

Among the emerging tools available to museums, artificial intelligence presents significant opportunities when implemented thoughtfully and ethically. One particularly promising application is the use of AI systems that enable automatic language translation and real-time interaction with digital personal tutors. These systems can adapt dynamically to the visitor's cultural background, language, and educational level, offering a high degree of personalization at a relatively low cost. For instance, AI can detect and interpret a visitor's spoken language through a microphone and then respond with both written and audio feedback in the same language. Furthermore, based on the complexity of the visitor's question, the system can infer whether it is interacting with a child, an adult, or even a subject-matter expert, adjusting its responses accordingly (this is precisely how the virtual museum presented and realized here interacts). Building on earlier experiments with AI-driven tutoring systems [29], this approach introduces a virtual tutor within a clearly defined ethical framework. The tutor is not intended to replace human educators or guides but rather to act as a flexible and adaptive support tool that enhances learning. This is particularly valuable for visitors with special educational needs, as it fosters inclusive and differentiated access to cultural content. While past technological innovations in museums often prioritized immersive experiences, such as virtual or augmented reality, current research is shifting toward AI solutions that emphasize personalization, accessibility, and user engagement. The integration of AI into museum education reinforces the institution's evolving role as an inclusive, participatory learning environment. However, as Carci, Caforio, and Gamper [30] (p. 283) point out, the effective use of digital technology in museums depends on strategic planning and well-defined implementation frameworks. These strategies must take into account the museum's specific starting conditions, the educational goals it aims to achieve, and the careful selection of technologies that align with long-term audience development objectives. Only through intentional, goal-oriented planning can AI serve not merely as a symbol of innovation, but as a meaningful instrument of transformation in the museum's educational mission.

5. Case Study 2: Designing a Multilingual NPC for a Virtual Museum Experience

As part of the educational project developed for the Cyprus Folk Art Museum, a Non-Playable Character (NPC) based on the collaboration of the University of Salerno with Convai technology (http://www.labh.it/disuff/cyprus-folk-art-museum/ accessed on 22 June 2025) was developed. This NPC has been integrated into the virtual museum environment, serving as a dynamic virtual tutor, which is present in a game environment realized with Unity3D (version 6000.0.43, Released on 18 mar 2025) that has a *fythkiotika* inside it in which details can be zoomed in on thanks to some selection buttons. The game can also be played inside a *MetaQuest* helmet. This NPC enhances the edugame experience by transforming it into a rich, personalized educational journey that bridges cultural heritage and contemporary learning methodologies, i.e., the transformative one presented above, which takes into account Deweyian pedagogical activism and the theme of constructivism in museum contexts. The NPC operates as an interactive guide, providing detailed explanations of exhibited items, especially the traditional Cypriot fythkiotika textiles, which as mentioned several times, are one of the most emblematic forms of Cypriot loom embroidery. The NPC is powered by a transformemetar-based large language model (LLM), fine-tuned on a curated dataset covering historical, pedagogical, and museum-related topics. Its architecture is based on recent advancements in natural language processing, enabling it to manage conversational flows, understand context, and respond in an adaptive, age-appropriate manner. Although the NPC does not engage in real-time online learning for privacy and ethical compliance, the system is periodically updated through offline supervised learning. Anonymous logs of user interactions are analyzed to improve the answer accuracy, conversational relevance, and engagement strategies. To ensure the historical and cultural accuracy of its responses, the NPC has been trained using a curated set of scholarly and historical sources specifically focused on fythkiotika. These include *The Fytiotika Textiles* by Pantazis [31], *Fabrics from Cyprus* by Papadimitriou [32], and *Paphos Embroideries* by Pieridis [33], as well as other key documents such as Dixon and McKee's *Fytiotika Weaving Project* and *Keeping the Thread* [34], Astreou-Christoforou's *Fytiotiko weaving* [35], Lewis's *A Lady's Impressions of Cyprus in 1893* [36], and de Mas Latrie's *History of the Island of Cyprus under the Rule of the Princes of the House of Lusignan* [37]. In addition to these study texts, some websites indicated by the Director of the Nicosia Folk Art Museum, UNESCO pages, and other official sites were used. These references provide the foundational knowledge the NPC draws upon when responding to user queries or guiding visitors through the historical evolution, technical characteristics, and cultural significance of *fythkiotika* textiles. The NPC, informed by these sources, conveys how *fythkiotika* developed as a distinct textile tradition with motifs, colors, and weaving techniques that set them apart from other forms of embroidery on the island. It also helps visitors appreciate the ongoing relevance of this practice, which continues to symbolize the richness of Cypriot folk heritage.

The AI-powered edugame is physically installed inside the Cyprus Folk Art Museum and can be experienced by visitors using either a MetaQuest VR headset or standard desktop equipment with a keyboard and mouse. In addition to the in-museum experience, the NPC can also be accessed remotely via the Cyprus Folk Art Museum's official website. This enables users to interact with the virtual museum and the AI tutor from any location with an Internet connection. The digital environment faithfully replicates a section of the museum space, allowing users to move freely and interactively within it. The experience is both exploratory and guided: visitors can choose their own paths or follow narrative prompts suggested by the system. The interaction with the multilingual Non-Playable Character (NPC) is enabled through voice and text, supporting multiple languages. A speech recognition module integrated into the system allows the NPC to detect the spoken

language through the user's microphone. The input is transcribed, semantically interpreted, and responded to through both synthesized speech and written text, ensuring an inclusive and multimodal interaction. The NPC, trained using curated historical and academic sources, can detect the user's language and respond accordingly, offering context-sensitive information adapted to the visitor's age and background. Users can approach virtual exhibits, click to reveal information, and ask spontaneous questions, engaging in a dialogical and personalized exploration of the cultural content.

This hybrid approach—combining immersive navigation, multilingual interaction, and pedagogically designed content—ensures accessibility, inclusion, and a deeper engagement with the intangible heritage of Cyprus. By embedding the experience within the museum itself, the project also fosters continuity between physical and digital cultural mediation.

Beyond its factual guidance, the NPC personalizes the user experience by adapting the content based on individual interests and prior interactions. It narrates stories linked to the objects on display, proposes educational activities and quizzes, and offers interactive support—such as adjusting visual settings for accessibility—to ensure an inclusive learning environment. Drawing from educational resources provided by human instructors, the NPC also reflects a coherent pedagogical approach, facilitating continuous, individualized learning both within and beyond the museum space. This virtual character transforms the digital museum visit into an experience that is at once culturally grounded, pedagogically sound, and technologically advanced, fostering a deeper engagement with Cyprus's intangible cultural heritage. In the context of a museum-based edugame, the implementation of a multilingual NPC is of fundamental importance to ensure accessibility, inclusion, and meaningful engagement for a diverse audience. Museums increasingly welcome visitors from various linguistic and cultural backgrounds, particularly in international or touristic contexts such as Cyprus. An NPC capable of communicating in multiple languages can serve as a bridge between the cultural content and the visitor, allowing individuals to fully understand and appreciate the significance of the artifacts on display, regardless of their native language. This linguistic flexibility not only enhances cognitive comprehension but also fosters a sense of welcome and belonging, reinforcing the museum's role as an inclusive educational space.

After a very recent meeting of research teams from the University of Salerno and the University of Nicosia, equally significant is the decision to redesign, as soon as possible, the NPC as a female avatar dressed in traditional Cypriot attire. This choice is grounded in both historical accuracy and cultural authenticity. Traditionally, the production of *fythkiotika* textiles was a female-dominated practice, carried out by women who not only wove the textiles but also preserved and transmitted the meanings embedded in each geometric pattern and color combination. By representing the NPC as a woman wearing the traditional costume of the region (see Figure 6), the game pays homage to the custodians of this intangible cultural heritage and affirms their role as bearers of knowledge and craft. Moreover, such a representation supports the narrative immersion, allowing players—particularly children and students—to engage with a credible and contextually appropriate figure who embodies the lived experience of the craft. This design decision not only strengthens the authenticity of the educational content but also supports gender-sensitive representation in digital heritage initiatives.

Figure 6. A possible design for a new NPC.

6. Conclusions

In today's social landscape, in light of the foregoing discussion and synthesizing the initial considerations inspired by Dewey's activism and the significance of narration [24], museums have evolved beyond their traditional role as historical object displays to become dynamic educational centers committed to making cultural heritage accessible to all, regardless of physical, economic, or social limitations. It is also a place where students can go for transformative learning experiences. Simultaneously, technology applied in this context serves as a bridge between the past and future, intriguing new generations when effectively utilized. This educational game is about to be installed within the Cyprus Folk Art Museum as part of an experimental initiative to enhance visitor engagement through innovative, culturally grounded digital experiences. A six-month data collection phase will be aimed at evaluating the inclusivity of the game's design, paying particular attention to its accessibility, cultural sensitivity, and educational effectiveness across diverse user profiles. The insights gathered during this period will inform potential refinements to the system, ensuring that the digital experience not only aligns with pedagogical and museological standards but also meets the needs of all visitors, including those with special educational needs or physical limitations. This process represents a crucial step in validating the role of intelligent, culturally informed NPCs within virtual museum environments as tools for inclusive and adaptive learning.

This paper presents a pilot study; therefore, it is not yet possible to draw definitive conclusions regarding the effectiveness of the AI-based tutor. However, this work builds

upon a series of previously tested and documented educational tools, the result of more than 10 years of work on Unity 3D models and edugames in education, developed by the Laboratory Teaching Learning Centre for Education and Inclusive Technologies, led by Elisa Frauenfelder. In particular, prior research has already investigated the use of head-mounted display (HMD) technologies to support teaching and learning through virtual learning environments in non-formal educational contexts (http://www.labh.it/disuff/ accessed on 22 June 2025). These technologies have been employed in both in-person and remote learning activities, such as those implemented at the Virtual Archaeological Museum of Herculaneum (https://www.museomav.it/?lang=en accessed on 22 June 2025). The design of these tools is grounded in the recognition that edugames can facilitate perspective-taking and that learning environments must be co-designed with learners—particularly those with special educational needs—based on their specific requirements. Additionally, the emotional and graphic impact of 3D engines and modeling technologies used in XR/VR/AR environments plays a critical role and must be carefully considered.

Author Contributions: Software, S.D.T.; writing—original draft preparation, M.D.T., E.P., A.F., A.M., L.C., F.P., and S.D.T.; supervision, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The software is available for download at the following link: http://www.labh.it/disuff/cyprus-folk-art-museum/ accessed on 22 June 2025, under the "Download" section, and may be used for independent trials and evaluations of the NPC (Non-Player Character) system. Additional data related to NPCs developed for museum environments—on which this adaptation for the Cyprus Folk Art Museum in Nicosia is based—has been published previously. The entire development process has been carried out within the "Teaching Learning Center for Education and Inclusive Technologies–Elisa Frauenfelder" laboratory, part of the Department of Human, Philosophical, and Educational Sciences at the University of Salerno. The lab conducts research in the field of Educational Technology. Its website features a range of research projects employing 3D printing, virtual reality, and augmented reality in educational and museum contexts, under the scientific direction of Professor Maurizio Sibilio. For further information or potential collaborations, interested parties are encouraged to contact the laboratory directly, which will be pleased to share additional insights and explore opportunities for cooperation.

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

References

1. Michalopoulos, S.; Xue, M.M. Folklore. *Q. J. Econ.* **2021**, *136*, 1993–2046. [CrossRef] [PubMed]
2. Alter-Muri, S.; Klein, L. Dissolving the Boundaries: Postmodern Art and Art Therapy. *Art Ther.* **2007**, *24*, 82–86. [CrossRef]
3. Dewey, J. Experience and Education. *Educ. Forum* **1986**, *50*, 241–252. [CrossRef]
4. Dewey, J. *Art as Experience*; A Putnam Capricorn Book: New York, NY, USA, 1958.
5. Sibilio, M. *La Semplessità: Proprietà e Principi per Agire il Cambiamento*; Morcelliana: Brescia, Italy, 2023.
6. Mörsch, C.; Sachs, A.; Sieber, T. (Eds.) *Contemporary Curating and Museum Education*, 1st ed.; Transcript Verlag: Bielefeld, Germany, 2017. Available online: https://www.jstor.org/stable/j.ctv371ck10 (accessed on 22 June 2025).
7. Mayer, M.M. Bridging the Theory-Practice Divide in Contemporary Art Museum Education. *Art Educ.* **2005**, *58*, 13–17. Available online: https://www.jstor.org/stable/27696060 (accessed on 22 June 2025). [CrossRef]
8. Black, G. *Transforming Museums in the Twenty-First Century*; Routledge: London, UK, 2012.
9. Scott, C. Museums: Impact and value. *Cult. Trends* **2006**, *15*, 45–75. [CrossRef]
10. Walker, S. *Teaching Meaning in Artmaking*; Davis: Worcester, MA, USA, 2001.
11. Deiter, R. The Use of Humor as a Teaching Tool in the College Classroom. *NACTA J.* **2003**, *20*, 20–28.
12. Gonizzi Barsanti, S.; Caruso, G.; Micoli, L.L.; Covarrubias Rodriguez, M.; Guidi, G. 3D Visualization of Cultural Heritage Artefacts with Virtual Reality devices. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci* **2015**, *XL-5/W7*, 165–172. [CrossRef]
13. Shah, N.; Ghazali, M.A. Systematic Review on Digital Technology for Enhancing User Experience in Museums. In *User Science and Engineering*; Abdullah, N., Wan Adnan, W., Foth, M., Eds.; Springer: Gateway East, Singapore, 2018; pp. 35–46.

14. Devitt, A.M. Editorial. *Mus. Int.* **2018**, *70*, 3–4. [CrossRef]
15. Chiscano, C.M.; Jiménez-Zarco, A. Towards an Inclusive Museum Management Strategy: An Exploratory Study of Consumption Experience in Visitors with Disabilities—The Case of the Cosmocaixa Science Museum. *Sustainability* **2021**, *13*, 660. [CrossRef]
16. Hutson, J.; Hutson, P. Museums and the Metaverse: Emerging Technologies to Promote Inclusivity and Engagement. *Fac. Scholarsh.* **2023**. Available online: https://digitalcommons.lindenwood.edu/faculty-research-papers/452 (accessed on 22 June 2025).
17. Dewey, J. *Arte Come Esperienza*; Aesthetica: Palermo, Italy, 2000.
18. Hein, G.E. *Learning in the Museum*; Routledge: London, UK, 2002.
19. D'Ambrosio, M. Quando il museo si fa patrimonio comune: Le donne del teatro di Forcella a Palazzo Zevallos. *MeTis. Mondi Educ. Temi Indagini Suggest.* **2017**, *7*, 468–503.
20. Bruner, J. Life as Narrative. *Soc. Res.* **1987**, *11*, 11–32. [CrossRef]
21. Clifford, J.; Bartezzaghi, S.; Bettini, M.; Bolzoni, L.; Dionigi, I.; Vegetti Finzi, S.; Zoja, L. *Narrare Humanum est: La Vita Come Intreccio di Storie e Immaginari*; UTET: Milano, Italy, 2023.
22. Hein, G.E. The Role of Museums in Society: Education and Social Action. *Curator* **2005**, *48*, 357. [CrossRef]
23. Stoškutė, N. Tension Between Everyday Practice and the New Museology Theory: A Case of the National Gallery of Art in Vilnius. *Art Hist. Crit./Meno Istor. Krit.* **2017**, *13*, 76–87. [CrossRef]
24. Hein, G.E. Museum Education. In *A Companion to Museum Studies*; Blackwell Publishing: Oxford, UK, 2006; pp. 340–352.
25. Kaplan, A.; Garner, J.K.; Rush, A.; Smith, M. Designing for Diverse Museum Visitors' Identity Exploration Around Inventiveness. *Front. Educ.* **2023**, *8*, 1078001. [CrossRef]
26. Guerra, L.; Caprara, B.; Scala, C. Virtual and Real: A Proposal for the Creation of an Integrated Museum. *Ricerche Pedagog. Didatt. J. Theor. Res. Educ.* **2008**, *3*. [CrossRef]
27. Wang, X.; Wang, W.; Wang, P.; Hu, H.; Li, C. A bibliometric analysis of immersive technology in museum exhibitions. *Front. Virtual Real.* **2023**, *4*, 1240562. [CrossRef]
28. Shehade, M.; Stylianou-Lambert, T. Virtual reality in museums: Exploring the experiences of museum professionals. *Appl. Sci.* **2020**, *10*, 4031. [CrossRef]
29. Coppola, F. A Salerno il tutor universitario è un avatar. *Rai—Radiotelev. Ital.* **2024**. Available online: https://www.rainews.it/tgr/piemonte/video/2024/03/avatar-ia-universita-salerno-071d1616-01af-4b34-b59d-562cbe40cf8e.html (accessed on 22 June 2025).
30. Carci, G.; Caforio, A.; Gamper, C. Digital Technologies and Museums: Augmented Reality, Learning and Audience Development. *Form@re Open J. Formaz. Rete* **2019**, *19*, 274–286. [CrossRef]
31. Pantazis, E. *The Fytiotika Textiles*; Ministry of Agriculture and Natural Resources: Krishi Bhavan, New Delhi, India, 1976.
32. Papadimitriou, E. *Fabrics from Cyprus*; En Typois: Nicosia, Cyprus, 2008.
33. Pieridis, A. Paphos Embroideries. In *Cypriot Folk Art*; Publications of the Society of Cypriot Studies: Nicosia, Cyprus, 1980; No. 6; pp. 3–20.
34. Dixon, S.; McKee, M. *Fytiotika Weaving Project, Short Report on an Exploratory Visit*; Printed document with limited circulation; Laona Foundation: Nicosia, Cyprus, 2010.
35. Astreou-Christoforou, J. *Fytiotiko Weaving*; Laona Foundaton Publication Series; Limassol, Cyprus, 2021. Available online: https://laona.org/fytiotiko-weaving-guide/ (accessed on 22 June 2025).
36. Lewis, A. *A Lady's Impressions of Cyprus in 1893*; Remington and Company Ltd.: London, UK, 1894.
37. Mas Latrie, L. *de Histoire de l'île de Chypre sous le règne des princes de la maison de Lusignan*; Imprimerie impériale: Paris, France, 1852; Volume III, pp. 1852–1855.

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.

computers

MDPI

Article

Exploring the Factors Influencing AI Adoption Intentions in Higher Education: An Integrated Model of DOI, TOE, and TAM

Rawan N. Abulail [1,*], Omar N. Badran [2], Mohammad A. Shkoukani [3] and Fandi Omeish [4]

[1] Department of Computer Science, Philadelphia University, Amman 19392, Jordan
[2] Department of Business Management, Istanbul Aydin University, Istanbul 34295, Türkiye; onidalbadran@stu.aydin.edu.tr
[3] Department of Computer Science, Applied Science Private University, Amman 11931, Jordan; m.shkokani@asu.edu.jo
[4] Department of E-Marketing and Social Media, Princess Sumaya University for Technology, Amman 11941, Jordan; f.omeish@psut.edu.jo
* Correspondence: rabulail@philadelphia.edu.jo

Abstract: This study investigates the primary technological and socio-environmental factors influencing the adoption intentions of AI-powered technology at the corporate level within higher education institutions. A conceptual model based on the Diffusion of Innovation Theory (DOI), the Technology–Organization–Environment (TOE), and the Technology Acceptance Model (TAM) combined framework were proposed and tested using data collected from 367 higher education students, faculty members, and employees. SPSS Amos 24 was used for CB-SEM to choose the best-fitting model, which proved more efficient than traditional multiple regression analysis to examine the relationships among the proposed constructs, ensuring model fit and statistical robustness. The findings reveal that Compatibility "C", Complexity "CX", User Interface "UX", Perceived Ease of Use "PEOU", User Satisfaction "US", Performance Expectation "PE", Artificial intelligence "AI" introducing new tools "AINT", AI Strategic Alignment "AIS", Availability of Resources "AVR", Technological Support "TS", and Facilitating Conditions "FC" significantly impact AI adoption intentions. At the same time, Competitive Pressure "COP" and Government Regulations "GOR" do not. Demographic factors, including major and years of experience, moderated these associations, and there were large differences across educational backgrounds and experience.

Keywords: AI adoption; diffusion of innovation theory (DOI); higher education; structural equation modeling (SEM); technology–organization–environment (TOE) framework; technology acceptance model (TAM)

1. Introduction

The embrace of advanced digital methods has become essential to today's business climate due to the constant demand for competition and efficiency. This has completely transformed the organization's work [1]. Companies increasingly utilize AI technologies to enhance productivity, decision-making, and efficiency [2]. The ubiquitous application of AI in marketing, education, manufacturing, and finance has shown positive performance and productivity outcomes because it transcends human cognitive constraints [3]. AI technologies are swiftly growing in organizations, disrupting traditional business processes and entering areas where human talent once thrived [4]. AI can cover a range of skills, from speech recognition to problem-solving and learning, that mimic human cognition [5]. Machine learning techniques cannot effectively leverage AI, preventing organizations from

extracting patterns and rules from extensive datasets. This limits organizations' ability to improve operational efficiency and make well-informed decisions [6]. Integrating AI technologies into companies offers considerable opportunities for enhancing business value chains, decision-making assistance, knowledge management, forecast maintenance, customer support, and relationship management [7]. Despite rising investments in AI technologies, organizations frequently encounter challenges in achieving the desired results. These problems include constrained budgets, skill deficiencies, and a lack of awareness, all of which hinder the widespread integration of AI [8]. Accordingly, there is an increasing necessity to identify the key factors influencing the successful implementation of AI technologies on an organizational scale [9]. Artificial Intelligence (AI) systems have emerged as a vital element in ensuring organizational effectiveness and improving student experiences in the ever-changing world of higher education [10]. The study reveals the technical and socio-environmental factors influencing the corporate-level adoption of AI-based technologies in higher education. It builds an abstract framework anchored in the DOI "theory of Everett Rogers on how new technology or ideas spread across society" [11] and the TOE model, "a model that highlights technological, organizational, and environmental aspects that influence an organization's technology adoption" [12]. The research also incorporates the Technology Acceptance Model (TAM) [13], which provides insight into users' behavioral intentions by focusing on perceived usefulness and perceived ease of use as the main drivers of technology acceptance [14]. This research aims to gain a deep understanding of the intricate processes underlying the use of AI in higher education. The paper examines the challenges to the implementation of AI in higher education. It explores "C", "CX", "UX", "PEOU", "US", "PE", "AINT", "AIS", "AVR", "TS", "FC", "COP", and "GOR". The research aims to uncover how these factors shape decision-making. It further evaluates the impact of age, gender, education, and experience. Understanding these interactions is crucial for AI adoption in higher education. The study provides insights into the complexities of AI adoption and aims to contribute to adoption strategies. It investigates the dynamics of AI adoption in higher education institutions and informs decision-making and planning for AI integration. Quantitative research employing a standardized questionnaire across three nations met the research objectives. "CB-SEM) was used in IBM SPSS AMOS version 24 to assess the associations of the study's latent variables. It was chosen for theory testing and model validation, especially with the DOI, TOE, and TAM models [15,16]. CB-SEM is preferred for verifying a theoretically grounded model in this study compared to (PLS-SEM), which is utilized for exploratory or predictive research. As Byrne (2016) states, CB-SEM allows simultaneous evaluation of measurement and structural models, including measurement error and indirect effects [17]. Global model fit indices offer a thorough and confirmatory approach to assessing structural relationships. (MLE) was used to estimate the model. The model fit was assessed using recognized indices, including CFI, TLI, RMSEA, and χ^2/df [18]. This analytical technique ensured the model's dependability, validity, and accuracy in hypothesis testing accuracy.

2. Literature Review

2.1. Higher Education

In higher education, curiosity towards using AI is diminishing. Making AI available to higher education has become all the more exciting because it can completely change the way that people learn and teach. Numerous studies have emphasized the profound impact on higher education, highlighting the importance of recognizing the drivers that drive AI use in the classroom. This question of AI in higher education is becoming more intriguing because it has the potential to reshape the way people teach and learn altogether. Greenhalgh et al. emphasized the need to understand the diffusion of innovations within service

organizations, particularly to pay attention to parameters such as communication channels and message delivery speed. This aligns with the research question of whether service organizations will use AI-based technologies [19]. Numerous studies have emphasized the profound influence AI has on higher education, underscoring the significance of comprehending the factors that affect the adoption of AI-based technologies in educational environments [20–23]. "C", "CX", "UX", "PEOU", "US", "PE", "AIS", "AVR", "COP", "GR", "TS", and "FC" and demographic variables such as age, gender, education, and years of experience all contribute to the intricate web of factors that influence the decision to adopt AI-based technologies [24]. Research discoveries have proven that incorporating AI in higher education is impacted by factors such as perceived risk, facilitating conditions, and expected effort. These factors, in contrast, do impact individuals' attitudes toward and intentions to utilize AI [25]. The likelihood of AI having a significant impact on higher education is highlighted in the literature, with expectations of considerable growth in AI implementation in the education sector [26]. As AI evolves, universities will need to adopt and leverage AI solutions for their educational activities. Recognizing the drivers of the use of AI in higher education and addressing the obstacles to its use are key to fully exploiting AI's educational potential.

2.2. Artificial Intelligence

AI is increasingly entering the field of higher education teaching and learning. Numerous studies have focused on AI and educational environments, specifically on factors that impact the adoption of AI-based technology.

Several key factors have been recognized as influencing the intention to adopt AI-based technologies within higher education. These include "C", "CX", "UX", "PEOU", "US", "PE", "AIS", "AVR", "COP", "GR", "TS", and "FC" [27]. Studies have highlighted AI's transformative potential within educational spheres, where breakthroughs like unobtrusive brain–computer interfaces, coupled with AI, pave the way for pioneering pedagogical methods [28].

2.3. Compatibility

Rogers defined 'compatibility' as "the extent to which an innovation is consistent with the values, practices, and needs of potential adopters." [11]. For AI-based technologies in higher education, alignment plays an important role in determining adoption strategies and institutional readiness [29]. AI solutions must adapt to prevailing educational methodologies and systems to become effective and widely adopted across educational facilities. The study underemphasizes compatibility as a crucial factor in adopting technological changes, such as cloud computing and e-commerce. These studies highlight that integrating new technologies with an organization's existing systems and practices is crucial for organizations to effectively leverage and benefit from technological advancements [29,30]. Similarly, in the AI applications for higher education, AI-driven tools must fit the context, curriculum requirements, and pedagogical practices. This consistency is crucial in order to ensure that educators and students adopt and effectively use these tools. Incorporating AI into learning environments is connected to several variables, including personalization, usability, and interactivity [24,31]. All of these factors play a key role in leveraging AI technologies for the differing needs and preferences of students and teachers. Further, openness to experimentation and compatibility between AI solutions and institutional goals and objectives are essential to the fit that motivates them.

The successful implementation of AI in education, as highlighted by [32]. Studies have also explored compatibility in the deployment of AI in healthcare, HR, and e-learning settings. It stresses the importance of perceived compatibility with organizational goals,

technological infrastructure, and user requirements [32,33]. Understanding factors influencing compatibility and addressing barriers to harmonization is key to making sure that AI technologies are used effectively across different organizational contexts. compatibility, in short, is an important factor in adopting AI technologies in universities and other fields. With AI solutions that are in sync with practices, values, and systems, organizations can make it easier for AI technologies to be used and integrated. resulting in increased efficiency, effectiveness, and creativity in educational processes. Therefore, we propose:

H1a: *There is a significant impact of "C" on AI adoption intentions.*

2.4. Complexity

The adoption of AI technologies in organizational contexts, such as higher education, is heavily dependent on their level of sophistication. In this context, 'complexity' refers to the perceived difficulty of using and implementing AI solutions in schools. Understanding the factors contributing to this complexity is key to the successful integration and maximization of AI technologies within higher education [23]. Research suggests that educational institutions struggle to implement AI technologies, which can negatively impact students' learning experiences. Furthermore, applying AI for teaching, student support, and other administrative tasks presents challenges for these institutions and requires further research to address these issues [23]. The deployment and execution of AI systems face hurdles due to their complex nature and the varying needs of educational stakeholders. The complexity of technology is not the only factor influencing the adoption of AI-based technologies; human elements are equally important. ALTakhayneh et al. explored the psychological resistance of teachers to digital innovation, highlighting the importance of overcoming psychological barriers and fostering positive attitudes toward educational technology to facilitate the adoption process. Navigating the intricacies of AI adoption in educational settings necessitates overcoming resistance and cultivating a welcoming attitude towards AI technologies. Moreover, the successful integration of AI-based technologies in higher education depends on the interplay of complexity with other elements such as "C", User Experience, and organizational preparedness. Addressing these multifaceted challenges and adopting strategies to enhance user acceptance and organizational support are crucial for educational institutions aiming to successfully navigate the complexities of AI adoption and unlock AI's transformative power in teaching and learning. In summary, grasping and addressing the complexities associated with adopting AI-based technologies is vital for their successful incorporation and use in higher education. Recognizing these challenges and implementing strategies to overcome obstacles will enable educational institutions to refine the adoption process and utilize AI technologies to improve educational outcomes [34]. Therefore, we propose:

H1b: *There is a significant impact of "CX" on AI adoption intentions.*

2.5. User Experience

User Experience "UX" plays a key role in utilizing AI technologies, particularly in the higher education sector. UX encompasses the overall user experience, satisfaction, and interaction with AI platforms and applications. Understanding the factors that influence UX is vital for enhancing user adoption, engagement, and the implementation of AI in educational environments [35,36]. Many studies emphasize the importance of UI design that addresses the diverse needs and preferences of users, especially those with limited reading skills. Numerous studies have underscored the critical nature of usability, accessibility, and user-centered design principles in developing intuitive, engaging, and inclusive artificial intelligence "AI" applications [35].

By creating intuitive user interfaces that are responsive to users' needs, colleges and universities can vastly improve the "UX" of AI technologies at the higher educational level. Furthermore, AI's entrance into the healthcare space has reaffirmed UX's role in fostering trust and acceptance among doctors and patients. Transparency, reliability, and ease of use all contribute to good user behavior and perceptions of AI in the medical environment [36]. With an emphasis on a transparent, trustworthy, and simple AI system design, healthcare providers can enhance UX and build user trust. The application of goal-setting theory in understanding user adoption intentions has been utilized for AI-enabled mobile applications. Researchers have delved into the impact of users' internal states on their behavior and attitudes toward AI-powered mobile applications, considering AI technologies' intelligent and anthropomorphic characteristics [37]. To design AI applications that resonate with user expectations and preferences, it is essential to grasp users' perceptions of AI technology and its influence on their goal-setting behavior. Moreover, the UX of AI-driven systems, like mobile fitness apps, has been scrutinized from a UX perspective. Employing goal-setting theory, researchers have explored how users' views on informational and emotional support influence their adoption and ongoing engagement with AI-powered applications [37]. Enhancing user engagement and satisfaction over the long term can be achieved by improving the UX of AI technologies with personalized and supportive elements. For higher education and other sectors to boost user acceptance and engagement, prioritizing UX in designing and deploying AI-based technologies is imperative. By concentrating on usability, accessibility, transparency, and personalization, organizations can refine the UX of AI systems, leading to enriched user experiences and heightened adoption and use of AI technologies. Therefore, we propose:

H1c: *There is a significant impact of "UX" on AI adoption intentions.*

2.6. Perceived Ease of Use

(PEOU) is a fundamental construct within the Technology Acceptance Model (TAM), defined as "the degree to which a person believes that using a particular system would be free from effort" [14]. Perceived Ease of Use (PEOU) greatly affects consumers' attitudes and behavioral intentions regarding the adoption of new technology. In the realm of mobile website utilization, perceived ease of use (PEOU) has been demonstrated to favorably influence user happiness, underscoring its significance in improving user experience and acceptance [38]. In e-learning contexts, perceived ease of use (PEOU) directly affects users' attitudes and indirectly shapes their behavioral intentions via perceived usefulness, highlighting its essential role in technology adoption [39]. Moreover, research on mobile learning has shown that perceived ease of use (PEOU) substantially influences both perceived usefulness and user attitudes, thereby affecting their intention to utilize these technologies [40]. These findings collectively underscore that when people regard a technology as user-friendly, they are more inclined to consider it beneficial and cultivate a positive attitude toward its adoption. Therefore, we propose:

H1d: *There is a significant impact "PEOU" on AI adoption intentions.*

2.7. User Satisfaction

User satisfaction plays a major role in ensuring the adoption and continued use of AI-based technologies, especially in higher education institutions. It refers to the aggregate happiness and good experiences people have in interacting with AI systems and resources. Understanding the elements that impact user satisfaction is essential for improving engagement, acceptance, and implementation of AI technologies in educational environments. Research has emphasized the significance of user satisfaction in various

AI-powered systems, including mental health chatbots and mobile banking applications. Studies indicate that factors such as usability, service quality, and anthropomorphism influence user contentment and their continued use of AI technologies [37,41]. User satisfaction is pivotal to the successful deployment and continued use of AI technologies within higher education. It pertains to how well users' expectations and actual experiences correspond with the effectiveness and advantages offered by AI systems. Recognizing the elements that impact user satisfaction is vital for fostering favorable user perceptions, involvement, and sustained use of AI technologies in academic environments [42,43]. Moreover, integrating AI into educational settings has underscored the importance of user satisfaction in fostering the effective implementation of technology. Elements like system performance, ease of operation, and perceived advantages have been recognized as crucial factors influencing user satisfaction and acceptance of AI-driven tools in education, both in teaching and administrative tasks [43]. By emphasizing user-centered design and features, educational establishments can improve user satisfaction and ease the seamless incorporation of AI technologies into educational activities. Moreover, the advancement of AI technology in education depends on feedback systems that capture user satisfaction and preferences. The research underscores the importance of user feedback, usability testing, and iterative design in enhancing user satisfaction and driving innovation in educational AI applications [37]. By integrating user feedback into the development of AI technologies, educational institutions can guarantee that AI systems meet user expectations and enhance learning outcomes positively. In conclusion, it is imperative to consider user satisfaction when designing, implementing, and testing AI technology for positive user experiences and easy adoption of AI in higher education. Through prioritizing factors that boost user satisfaction, such as perceived usefulness, expectation-fulfillment, and individualized experience, education providers can enhance user interaction and the use of AI technology for better learning outcomes. Therefore, we propose:

H1e: *There is a significant impact of "US" on AI adoption intentions.*

2.8. Performance Expectation

Performance Expectation "PE" is an essential factor influencing the adoption intentions of AI-based technologies in different sectors, including higher education. The research conducted by [44] demonstrated that performance expectancy significantly and positively impacts behavioral intention, among other factors. This highlights the importance of PE in shaping individuals' attitudes towards adopting new technologies. Similarly, Rasheed et al. classified the existing literature on AI adoption into drivers and barriers, emphasizing the role of PE in influencing adoption decisions [45]. In examining e-learning adoption among students, Tarhini et al. highlight the significance of factors influencing student adoption behaviors, such as performance expectations. Their research enriches academic discourse by incorporating a range of variables that the University Information Disclosure System correlates with perceived performance outcomes, emphasizing the importance of fulfilling information needs to enhance enterprise agility. This study underscores the role of PE in influencing users' perceptions of technology performance within universities and evaluating them within the context of UK universities, offering significant perspectives on the influence of PE in students' acceptance of technology [46]. Lee et al. investigated what methods of satisfaction with the systems, highlighting the need for aligning technology with user expectations [47]. They also investigated the influence of quality library information resources on the satisfaction of postgraduate students at the Ignatius Ajuru University of Education library. The study emphasized the significance of pertinent information resources and their accessibility in boosting user satisfaction. It stressed the crucial role of the physical

environment in determining users' contentment with library services, pointing out the necessity for sufficient resources to fulfill user expectations [47]. Therefore, we propose:

H1f: *There is a significant impact of "PE" on AI adoption intentions.*

2.9. Introducing AI New Tools

Implementing new AI technologies in different organizational contexts, including universities, is impacted by a range of technological and socio-environmental factors. A study conducted by Henke examined university communication tactics and viewpoints on generative AI tools. The findings revealed significant variations in adopting AI tools among universities, which can be attributed to disparities in their approaches. The variation highlights the different methods and tactics used by universities to incorporate and utilize AI capabilities in their teaching settings [48]. In their study, Okunlaya et al. presented a novel conceptual framework that explores the use of AI library services to alter university education digitally. They highlighted the need to adopt AI technology to improve service delivery and encourage new behaviors in educational institutions [49]. The study conducted by Dora et al. revealed key factors that play a crucial role in the adoption of artificial intelligence in food supply chains. These factors include technology readiness, security, privacy, customer satisfaction, demand volatility, regulatory compliance, competitor pressure, and information sharing among partners. The study highlights the importance of these factors in promoting the adoption of AI. These aspects are essential in influencing the adoption of AI technology in many areas, such as education [42]. In addition, Gupta and Gupta highlight the combined effectiveness of need-based and curiosity-based experimentation in the adoption of AI technology for libraries. This study offers a comprehensive approach to controlling the adoption of AI technology in the educational scope [50]. Sallu, S.et al. focus on the utilization of AI in higher education. The study provides valuable information on how AI technologies can bring about significant changes and improvements in academic settings [51]. Saidakhror examines the influence of artificial intelligence on higher education and the economic aspects of information technology, demonstrating the various uses of AI tools in educational environments. The implementation of AI technologies in higher education institutions is impacted by various elements, such as technology preparedness, organizational backing, and individual perspectives. Comprehending these factors is crucial for universities to negotiate the intricacies of AI implementation, improve teaching and learning methods, and stimulate innovation in educational environments [52]. Therefore, we propose:

H1g: *There is a significant impact of "AINT" on AI adoption intentions.*

2.10. AI Strategic Alignment

AI strategic alignment "AIS" is one of the biggest drivers for the use and deployment of AI within higher education, particularly universities. Many researchers have captured the strategic implications of AI adoption and how AI activities align with institutional objectives and goals in the university context, which further indicates the importance of AIS in ensuring successful technology integration. Jarrahi et al. discussed in greater depth the strategic benefits of AI and its profound influence on the improvement of organizational learning. The study indicates the significance of AIS in supporting innovation in organizations and the effective operation of knowledge assets. In its attention to the slow, pathway-driven adoption of AI in organizational life cycles and continuous learning strategies, this study accentuates the strategic implications of AIS in navigating organizational successes towards long-term success [53]. Okunlaya et al. developed a new theoretical model for digitizing university teaching using AI library services. The study focuses on

the importance of adopting AI in the context of university library services to support educational outcomes and customer experiences, as well as the significance of AI systems in driving digital transformation efforts. According to the literature review, the strategic coherence of AI is crucial for its successful deployment and incorporation in higher education. Understanding the strategic implications of AI adoption and aligning AI initiatives with university objectives is essential for universities to successfully leverage AI technology and encourage innovation in the classroom [49].Therefore, we propose:

H1h: *There is a significant impact of "AIS" on AI adoption intentions.*

2.11. Availability of Resources

The Availability of Resources "AVR" is essential for successfully implementing and integrating AI-based technologies in higher education institutions, including universities. Multiple studies have examined the necessity of having enough resources to facilitate the adoption of technology and innovation in universities, particularly concerning AI efforts. Bearman and Ajjawi examined the educational significance of artificial intelligence in education, highlighting educators' need to modify their instructional approaches to integrate AI technologies proficiently. This study emphasizes the importance of having access to resources, such as AI tools and educational materials, in equipping students for a future that involves AI technology. It highlights how resources play a crucial role in influencing teaching methods in higher education [54]. The study specifically examined the factors that influence the employability of these graduates. This study emphasizes the significance of resource availability, such as high-quality education and skills development programs, in improving graduates' capacity to find employment and achieve success in the job market. It underscores the value of resources in supporting student outcomes [55]. Boonsiritomachai et al. emphasized the significance of technical resources in influencing the acceptance of modern technologies. They underlined the relevance of physical assets, such as networking, data, and computer hardware, in promoting the adoption of technology. This study emphasizes the importance of collaborative resources in establishing a scalable and adaptable basis for business applications, emphasizing the crucial role of technical resources in facilitating the integration of technology [56].

Availability of Resources "AVR" is a crucial element in facilitating the effective implementation and assimilation of AI-based technologies in higher education institutions. Universities must have sufficient resources and match their strategies with organizational goals in order to effectively utilize AI technologies and promote innovation in the education industry. Therefore, we propose:

H1i: *There is a significant impact of "AVR" on AI adoption intentions.*

2.12. Competitive Pressure "COP"

Competitive pressure "COP" influences the willingness of higher education institutions, especially universities, to adopt AI-based technology. Multiple studies have emphasized the significance of competitive pressure in promoting the adoption of technology and innovation in universities. These studies show the strategic necessity of effectively responding to competitive market dynamics. Hungund and Mani emphasized that organizational decisions about technology adoption are influenced by the external environment in which the company operates. They emphasized the strategic necessity of competitive pressure for firms to remain competitive in highly competitive markets. This research highlights the importance of competitive pressure in driving businesses to adopt innovative technologies to increase performance and increase their likelihood of survival in competitive environments [57]. Alsheibani et al. identified competitive pressure as a driving force for

the spread of innovative technologies, highlighting its significance in allowing companies to compete efficiently and preserve their competitive edge in the market. This research emphasized that companies operating in high-pressure situations tend to embrace innovative technologies to enhance their performance and enhance their likelihood of survival, emphasizing the strategic significance of addressing competitive challenges through technology adoption [58]. As Porter and Millar also pointed out, IT innovation can change the organization of industries, modify rules of competition, and reshape the landscape of competition, thus revealing the game-changing effect of technology on competition. The study emphasized that firms could take a step ahead in utilizing innovative technology and AI to improve their services and differentiate themselves from competitors in the market. This research illustrates how "COP" has an enormous impact on AI adoption intentions in higher education. In order to remain competitive, universities must adopt and innovate with technology to improve their performance and drive innovation in the education sector [59].Therefore, we propose:

H1j: *There is a significant impact of "COP" on AI adoption intentions.*

2.13. Government Regulations "GOR"

Government regulations "GOR" are essential in influencing the willingness of higher education institutions, namely universities, to adopt AI-based technologies. Multiple studies have emphasized the importance of government laws and regulations in promoting IT breakthroughs, such as integrating AI technology, in university environments. Alghamdi et al. (2020); Dora et al. (2021); Pan et al. (2022) [42,60,61] highlighted the crucial significance of governmental policies and laws in incentivizing the use of AI technology. These studies highlighted the support provided by state authorities in encouraging the implementation of AI technology through regulatory frameworks and guidelines, demonstrating the government's commitment to advancing technological progress in various industries, including higher education. Wong and Yap highlight the government's significance in shaping the adoption of artificial intelligence in accounting within micro, small, and medium firms. This exemplifies the extensive scope of government rules in influencing the deployment of AI technologies in many business contexts [62]. In addition, Dora et al. emphasizes that government regulators can exert pressure on companies to implement innovative technologies in their supply chain operations, demonstrating the impact of government regulations on technological progress in organizational procedures [42]. Ghani et al. highlights the essential role of government assistance in promoting the adoption of AI, underscoring the importance of government rules in influencing the acceptance of technology [63]. Within the domain of responsible AI governance, Alexander et al. (2024) [64] proposes that although governments are beginning to enforce formal regulatory measures to oversee AI, there is a requirement for ethics frameworks and "soft law" tools to supplement these rules. Moon proposes the notion of AI participatory governance, which involves several stakeholders, including government regulators, to ensure the fair and constructive utilization of AI for society's progress [65]. Farida et al. examine the impact of government AI systems on obtaining accountable results, emphasizing the crucial role of government rules in guaranteeing responsible development and implementation of AI technology [66]. In their study, Noordt and Misuraca put out a comprehensive framework that aims to enhance the effective implementation of AI systems in government settings. They highlight the various elements that play a role in AI adoption, going beyond just the technical components. "GOR" significantly impacts the willingness of organizations, particularly higher education institutions, to use AI-based technologies. The interplay of regulatory frameworks, ethical considerations, and stakeholder involvement highlights the intricate nature of AI governance, where responsible practices are essential

for optimizing the advantages of AI technologies while mitigating associated risks. An intricate comprehension of government regulations and their consequences is crucial for promoting innovation, guaranteeing adherence, and propelling sustainable technological advancement [67]. Therefore, we propose:

H1k: *There is a significant impact of "GOR" on AI adoption intentions.*

2.14. Technological Support

Technological support also determines the chances that higher education institutions will adopt AI technologies. Several studies have highlighted the importance of technological support for artificial intelligence adoption in the educational setting. Multiple studies have emphasized the need for technological assistance to incorporate artificial intelligence in educational environments. Hannan highlights the revolutionary influence of technological breakthroughs in higher education operations by demonstrating successful implementations of AI technologies in improving student learning experiences, student support services, and enrollment administration systems within educational institutions [26]. According to Greiner et al., the Technology Acceptance Model (TAM) and the Four-Sides Communication Model can be employed to understand how humans interact with AI and the adjustments needed for acceptance, specifically in the context of grading dissertations. This emphasizes the significance of matching technological assistance with the needs and expectations of users [55]. Crompton and Song explore how AI benefits students and faculty in higher education. They highlight how AI enables personalized learning, intelligent tutoring systems, facilitates collaboration, and automates grading. Such instances demonstrate how technological support improves educational processes [21]. Mohsin et al. highlight various factors that contribute to creating a favorable environment for adopting AI. These factors include sufficient funding, technological infrastructure, IT support, training programs, and institutional policies that promote the integration of AI. The author emphasizes the diverse range of technological support required to facilitate the adoption of AI [68]. In another study, an analysis was conducted on faculty attitudes, technology readiness, curriculum reform needs, and policy implications in the integration of AI in information technology education [69]. The study highlights the intricate nature of providing technological support in educational environments. Opesemowo and Adekomaya examine the qualitative elements of utilizing AI to promote Sustainable Development Goals in South Africa's higher education system. The study highlights the need for technological assistance in promoting sustainable educational practices [70]. Polyportis, A. performs a long-term investigation on the acceptance of AI, providing practical advice for AI developers and educational institutions to enhance student involvement with AI technologies, highlighting the ever-changing nature of technology assistance in educational settings. The presence of technology support plays a crucial role in determining the likelihood of AI-based technologies being adopted in higher education. The analyzed papers highlight the numerous ways in which technological support enhances educational processes, ranging from enhancing student learning experiences to optimizing administrative procedures. Comprehending the complex and diverse aspects of technology assistance is crucial for higher education institutions to successfully incorporate AI technologies, enhance student involvement, and foster innovation in instructional methods [71]. Therefore, we propose:

H1l: *There is a significant impact of "TS" on AI adoption intentions.*

2.15. Facilitating Conditions

Facilitating conditions have a critical role in shaping the likelihood of AI-based technologies being adopted in different organizational settings, such as higher education

institutions. According to Eftimov and Kitanoviki, facilitating conditions refer to the conducive surroundings and incentives that empower individuals to acknowledge the advantages of using AI technologies [72]. Tanantong and Wongras establish a connection between enabling situations and individuals' perceptions of the resources and support required for various behaviors [73]. Jain et al. highlight that the ease of conditions relies on users' perceptions regarding the accessibility of assistance and resources for utilizing technology within companies [22]. Mohsin et al. stress the positive effect of accommodating environments on effort expectancy, suggesting that a comfortable setting increases users' willingness to be able to effectively engage with AI systems [68]. Morrison examines barriers and facilitators for AI deployment in clinical settings in the NHS. Facilitating factors play a significant role in shaping the adoption strategies of AI technologies across multiple organizational settings. Creating optimal conditions, setting priorities, and creating the right environment are essential to using technology. Understanding and adapting to appropriate environments is essential for businesses, in particular higher education institutions, to be able to integrate AI, enhance the user experience, and drive creativity into adoption processes [74]. Therefore, we propose:

H1m: *There is a significant impact of "FC" on AI adoption intentions.*

2.16. AI Adoption Intention

Factors driving the decision to bring AI into universities include technological innovation, social-environmental aspects, and institutional factors. According to Chen et al. explored the conditions leading to the successful application of AI in China's telecommunications sector. The results of their study offer valuable insights for companies on how to choose and spend resources in adopting AI. The research pointed to the importance of understanding the factors leading to the successful adoption of AI in a particular industry context [75]. Furthermore, Chen et al. investigated the determinants impacting the acceptance of big data analytics and artificial intelligence in connection to operational effectiveness. The study highlighted the correlation between operational performance and environmental performance [76]. Furthermore, Bughin investigated the impact of a company's AI strategy on employment growth, specifically analyzing the strategic goals and resources affected by the use of AI. The study provided valuable insights into the consequences of using artificial intelligence on internal operations and various enterprise resources [77]. Govindan highlighted the significance of artificial intelligence in advancing sustainable and economically efficient innovation. He emphasized the importance of vendors integrating AI-based processes in order to adopt these technologies [78] effectively. Horani et al. (2023) provide valuable insights into the key elements that influence the implementation of artificial intelligence within organizations. By incorporating these findings into the discourse on the desire to embrace AI, companies, especially higher education institutions, can gain a comprehensive understanding of the key elements that impact decisions related to the adoption of technology [79]. Subsequently, individuals can utilize this acquired understanding to proficiently maneuver through the procedure of integrating AI and attain triumph. In summary, the choice to use AI-driven technologies in higher education institutions is influenced by factors such as technological support, favorable conditions, and regulatory frameworks. Understanding these attributes is crucial for businesses to successfully manage the use of artificial intelligence and foster innovation for enduring and sustainable expansion. Therefore, we propose:

3. Objectives

It seeks to identify and assess the key technological and socio-economic factors that may influence universities' decisions to adopt AI technologies. The research aims to understand the factors related to the intent to implement AI technologies, such as "C", "CX", User Experience, Perceived Usefulness, "PEOU", "US", "PE", "AIS", available resources, "COP", Government Policy, Technological Assistance, and Facilitating Environments.

Moreover, the study aims to analyze how demographic characteristics such as age, gender, education, and years of experience affect the relationships between the aforementioned parameters and the intention to use AI-based technology. Additionally, the study aims to develop a comprehensive understanding of how personal attributes are likely to interact with the tool and their implications for understanding the factors that govern decisions about AI implementation and its impact on organizational transformation.

4. Methodology

4.1. Participants

The study collected data from a diverse panel of 500 respondents, including university students and professional workers in higher education. This ensures that the contexts under which they might use AI-based technologies in their professional lives or higher education settings are considered. The viewpoints of learners and practitioners also help generate a comprehensive understanding of the intentions an individual might have to adopt AI use in educational settings.

4.2. Data Collection

The main tool for gathering data was a structured questionnaire that was created using well-known models, the (TAM), (DOI), and the (TOE) framework. The questionnaire was broken into multiple sections to evaluate different influencing factors and contained both demographic and construct-related items.

Instruments and constructs measuring the following were included in the survey:

The questionnaire was developed to assess constructs based on the DOI, TOE, and TAM models. Every construct was assessed using several indicators modified from recent validated studies. The survey included items related to compatibility (C1–C5), complexity (CX1–CX4), user experience (UX1–UX4), perceived ease of use (PEOU1–PEOU3), user satisfaction (US1–US6), performance expectation (PE1–PE3), AI strategic alignment (AIS1–AIS3), availability of resources (AVR1–AVR3), competitive pressure (COP1–COP3), government regulations (GOR1–GOR3), technological support (TS1–TS3), facilitating conditions (FC1–FC3), and AI adoption intentions (AIA1–AIA3).

Each item was scored using a 5-point Likert scale, with 1 denoting 'strongly disagree' and 5 denoting 'strongly agree.' The items were adapted and contextualized for the higher education environment. The full list of item codes and their source citations is provided in Appendix A. Open-ended questions were also incorporated to supplement the quantitative data in order to gather more information and set the statistical results in context.

4.3. Research Questions

How do the factors influence the adoption intentions of AI-based technologies in higher education institutions, aligned with the (TAM) and the (TOE) framework? Additionally, what are the mediating roles of demographic variables in these relationships?

The impact of key factors "C", "CX", User Experience "UX", "PEOU", "US", "PE", "AINT", "AIS", "AVR", "COP", "GOR", "TS", and "FC" together on AI adoption intentions among higher education students and faculty members in Türkiye, Canada, and the USA

is investigated. Additionally, the study aims to understand the moderating roles of age, gender, and years of experience in this relationship.

4.4. Deductive Approach

The current investigation utilized a deductive methodology. The deductive approach is typically initiated by formulating a hypothesis and subjecting it to careful observations or data collection. In the context of this particular inquiry, the investigation was initiated with a theoretical framework relating to the influence of the key factors on AI adoption intentions and then tested this theoretical perspective using empirical data from higher education students and faculty members.

4.5. Population

The study's participants were based in the USA, Canada, and Türkiye. Given that they are experts who play important roles in higher education, such as faculty members or student services specialists, they were specifically chosen to offer valuable insights into the factors influencing AI adoption intentions. Academic contacts, professional networks, and university directories were used to find participants. In Türkiye, a combination of Google Forms shared through institutional emails and paper-based questionnaires given out in person was used for recruitment. The Google Forms survey was disseminated to participants in Canada and the USA via direct messaging, sharing in pertinent academic groups, and professional social media sites like LinkedIn. Follow-up reminders were sent to encourage higher response rates, and participation was anonymous and voluntary. To investigate the relationships between variables and analyze the hypotheses put forth, the analysis relied on quantitative methods, covariance-based structural equation modeling (CB-SEM), and descriptive statistics. Multi-group comparisons and interaction effects within the SEM framework were also used to examine the moderating effects of demographic variables, including age, gender, years of experience, and educational background.

4.6. Sample Size Calculation

The sample size of the study of 367 university students and workers can be estimated using the sample size formula for a quantitative variable. This approach is often applied to prevalence or cross-sectional studies to make sure the sample size is large enough to provide statistical power and allow for the proper rejection of the data the null hypothesis if needed. G*Power 3.1, a popular statistical power analysis program for structural equation modeling (SEM), notably covariance-based SEM, was utilized to calculate the sample size for this investigation. G*Power efficiently estimates sample sizes using statistical test types, effect sizes, significance levels (α), and statistical power ($1-\beta$), providing trustworthy model estimation. This study utilized a priori power analysis for linear multiple regression (fixed model, R^2 departure from zero), a typical SEM sample size estimation proxy. These parameters were used:

Effect size (f^2) = 0.15 (medium, per Cohen's standards). Level of significance: $\alpha = 0.05$ power = $1-\beta = 0.95$

The proposed model has 12 predictors based on latent variables. These characteristics led G*Power to recommend 184 samples. The study has 367 participants, which exceeds this threshold and provides enough statistical power to detect relevant effects and yield reliable SEM results. G*Power follows SEM research best practices, where sample size is crucial for model fit, parameter stability, and generalizability [80].

4.7. Sampling Method

The study recruited 367 university students, faculty, and administrative staff through purposive sampling to investigate AI technology adoption in higher education. Researchers

employed a non-probability sampling approach to include only participants with relevant knowledge or experience in educational technology systems in their study.

4.8. Rationale for Purposive Sampling

Purposive sampling was used to gather specific and valuable information from individuals who use technology in education or are impacted by it. Participants were selected based on characteristics aligned with the study's focus, such as prior experience using digital platforms, decision-making roles in technology integration, or involvement in institutional technology planning. This strategy is particularly suitable for research exploring adoption behavior, as it prioritizes the inclusion of informed individuals who can contribute to a more nuanced appraisal of AI acceptance in universities [81].

Data were collected through an online survey distributed via institutional mailing lists, professional academic networks (LinkedIn, ResearchGate), and university forums across institutions in Türkiye, Canada, and the United States. Before participating, we informed all respondents of the study's objectives and obtained their informed consent.

4.9. TAM Construct Modification

Several fundamental concepts from the TAM model, originally created by Davis in 1989 and later refined by Venkatesh and Davis in 2000, were modified for this study. The conceptual framework was employed to reflect users' evaluations of AI-based technology in higher education using "PEOU" and "PU", with PU operationalized as "PE". In various organizational contexts, these constructs have consistently demonstrated their value in explaining people's intentions to adopt new technologies [14,82]. In line with extended (TAM) model frameworks, the inclusion of "US" and "TS" emphasizes the importance of post-adoption satisfaction and institutional infrastructure as factors influencing continued usage intention [83,84].

4.10. Theoretical Background and Hypothesis Structuring

Camisón and Villar-López defined innovation adoption as "the implementation of a new or significantly improved product (goods or services) or process, a new marketing method, or a new organizational method in business practices, workplace organization, or external relations." In other words, innovation organizational adoption refers to a firm's explicit choice to either adopt or use an innovative technology [85]. AI is an advanced and innovative technological domain [86]. Applying existing technology adoption models to study the implementation of AI at the organizational level poses significant challenges. AI encompasses the overall aspects of organizations, including their processes, data, talents, strategies, and structures [8]. Therefore, to examine the perspective of higher education organizations' implementation of AI, this article utilizes three frameworks: DOI [11], TAM [14], and the TOE framework [12]. DOI, or Diffusion of Innovation, is the process by which innovation or technology is shared among members of a social group over time through certain channels of communication [87]. The proposition suggests that the spread of a new idea or technology is determined by how people perceive it and the specific features of the technology itself. Diffusion is the process through which businesses, individuals, communities, or subsystems acquire and fully embrace innovative concepts, such as innovative technology, to advance in science and education. TAM does not incorporate a social element. Although Unified Theory of Acceptance and Use of Technology (UTAUT) includes social impact as a primary element in its model, it does not include the attitude variable. Attitude significantly determines the behavioral intention to use a specific technology, as highlighted in education [88]. The DOI theory and the TOE framework share significant commonalities, as noted by Baker (2012). The organizational and technological aspects of the TOE framework correspond to the innovative charac-

teristics and organizational context of the DOI model, respectively [89]. However, there are notable distinctions between DOI and TOE [88,90]. Unlike the DOI, the TOE system does not account for specific characteristics of individuals. Conversely, unlike the TOE model, the DOI hypothesis ignores environmental influences. Integrating the TOE and TAM elements is seen as essential to overcome DOI theory limitations in the context of technology adoption across multiple settings [91–93]. Therefore, by aligning the DOI, TOE, and TAM into a single framework, we can investigate the external and internal drivers behind an organization's adoption of AI. This means that the triad DOI-TOE-TAM model (Figure 1) is suitable for explaining the technological and socio-environmental aspects of AI implementation in organizations.

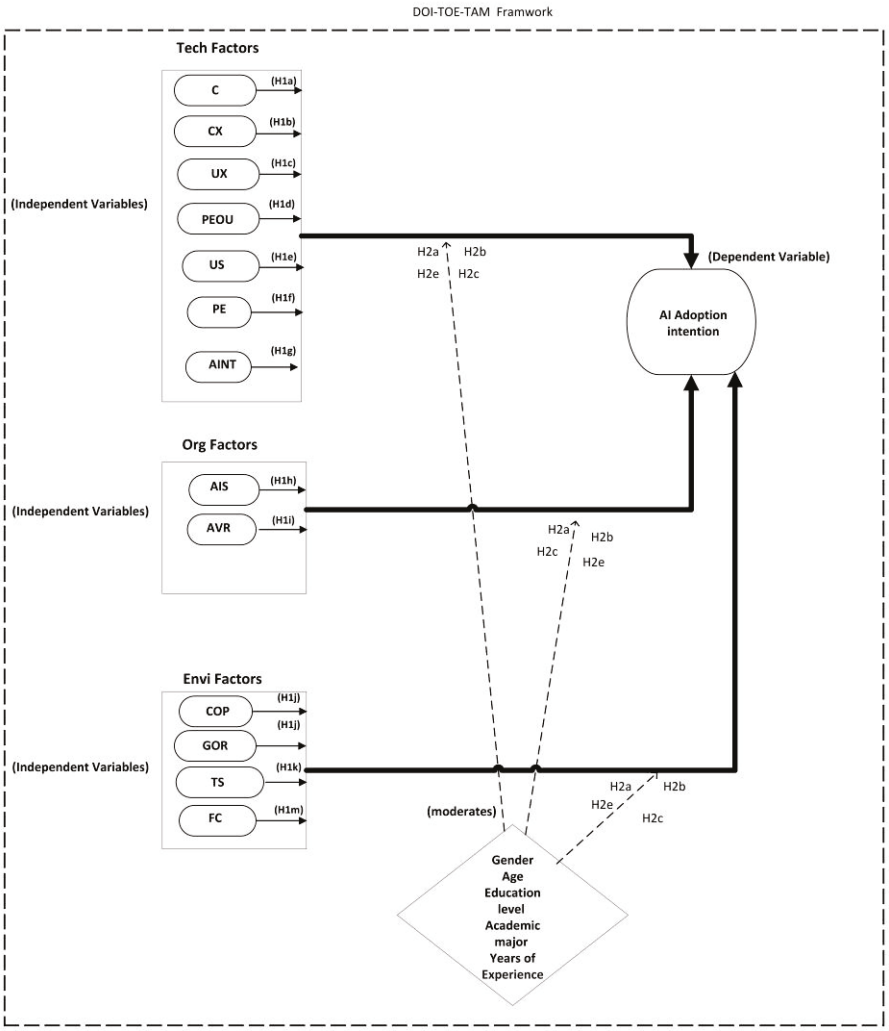


Figure 1. The study model.

4.11. Research Hypotheses

H1: *There is a statistically significant impact of key factors "C", "CX", "UX", "PEOU", "US", "PE", "AINT", "AIS", "AVR", "COP", "GOR", "TS", and "FC") together on AI adoption intentions at* $p \leq 0.05$.

This hypothesis was divided into 13 sub-hypotheses:

H1a: *There is a statistically significant impact of "C" on AI adoption intentions at* $p \leq 0.05$.

H1b: *There is a statistically significant impact of "CX" on AI adoption intentions at $p \leq 0.05$.*

H1c: *There is a statistically significant impact of "UX" on AI adoption intentions at $p \leq 0.05$.*

H1d: *There is a statistically significant impact of perceived ease of "PEOU" on AI adoption intentions at $p \leq 0.05$.*

H1e: *There is a statistically significant impact of US on AI adoption intentions at $p \leq 0.05$.*

H1f: *There is a statistically significant impact of "PE" on AI adoption intentions at $p \leq 0.05$.*

H1g: *There is a statistically significant impact of "AINT" on AI adoption intentions at $p \leq 0.05$.*

H12h: *There is a statistically significant impact of "AIS" on AI adoption intentions at $p \leq 0.05$.*

H1i: *There is a statistically significant impact of "AVR" on AI adoption intentions at $p \leq 0.05$.*

H1j: *There is a statistically significant impact of "COP" on AI adoption intentions at $p \leq 0.05$.*

H1k: *There is a statistically significant impact of "GOR" on AI adoption intentions at $p \leq 0.05$.*

H1l: *There is a statistically significant impact of "TS" on AI adoption intentions at $p \leq 0.05$.*

H1m: *There is a statistically significant impact of "FC" on AI adoption intentions at $p \leq 0.05$.*

4.12. The Research Model

Figure 1 shows the study model, which includes the independent variables (technological factors, organizational factors, environmental factors) and the dependent variable, which is AI adoption intention.

4.13. Data Analysis

A descriptive statistical analysis was performed to calculate the values, and both simple and multiple regression analyses were conducted to examine the relationships between the external variables and other study elements. The acquired data was analyzed, and the hypotheses of the study were evaluated using SPSS® Amos version 24 (IBM Corp., Armonk, NY, USA).

4.14. Descriptive Analysis

4.14.1. Sample Characteristics

Table 1 presents the demographic variables of the study sample; the male respondents were 51%, and the female respondents were 49%. The majority of respondents were between 34 and 44 years old (47.7%). The majority of respondents' education level was a PhD degree (59.2%). The majority of respondents' educational major was IT (37.1%), while respondents with other majors (32.7%) were in languages (16), structural design (seven), engineering (22), education (21), biology (five), economics (13), science (18), marketing (nine), MIS (five), and finance (four). The majority of respondents' work experience (44.7%) was 10 years and above, with (25.6%) of respondents using AI tools or apps for less than 6 months, (15%) of respondents using AI tools or apps from 6 months–less than 1 year, (13.1%) of respondents using AI tools or apps from 1 year–less than 2 years, and (46.3%) of respondents using AI tools or Apps for 2 years and more. The majority of respondents preferred Windows PC operating system to use their preferred AI tool, with a percentage of (73.3%), while the other respondents used Linux operating system.

Table 1. Sample Demographic Information.

Variable	Category	Count	Percent
Gender	Male	187	51
	Female	180	49
	Other	-	-
	Total	367	100
Age	18–24	55	15
	25–33	71	19.3
	34–44	175	47.7
	45–54	40	10.9
	55–65	26	7.1
	66 and older	-	-
	Total	367	100
Residence	Türkiye	71	19.4
	USA	192	52.3
	Canada	104	28.3
	Total	367	100
Education	Diploma degree	-	-
	Bachelor's degree	46	12.5
	Master's degree	104	28.3
	PhD	217	59.2
	Total	367	100
Educational Major	IT	136	37.1
	Management	74	20.1
	Accounting	4	1.1
	Medicine	22	6
	Pharmaceutical	11	3
	Other	120	32.7
	Total	367	100
Work Experience	Less than 2 years	58	15.8
	2 years–less than 6 years	78	21.3
	6 years–less than 8 years	25	6.8
	8 years–less than 10 years	42	11.4
	10 years and above	164	44.7
	Total	367	100
How long have you been using AI tools or apps?	Less than 6 months	94	25.6
	6 months–less than 1 year	55	15
	1 year–less than 2 years	48	13.1
	2 years and more	170	46.3
	Total	367	100
Where do you most use your preferred AI tool (which type of operating system do you use)?	Windows PC	269	73.3
	Mac OS (Mac Book)	24	6.5
	Android (Samsung, Sony, HTC, LG, Motorola. . .etc.)	19	5.2
	iOS (iPhone)	46	12.5
	Tablet	2	0.5
	Other	7	1.9
	Total	367	100

4.14.2. What Type of AI Tools Do You Use for Your Work or School Needs?

Table 2 shows that (74.7%) of all respondents used the ChatGPT-4 tool in their work or school needs, (36.8%) of all respondents used the QuillBot tool in their work or school needs, (67.6%) of all respondents used the Grammarly tool in their work or school needs, (9.8%) of all respondents used the Scholarcy tool in their work or school needs, (11.7%) of all respondents used the Scite tool in their work or school needs, (18.5%) of all respondents used the pdf.ai tool in their work or school needs, and finally (6.5%) of all respondents used other tools in their work or school needs which included Scispace AI, Tome AI, Cognigy.ai, Copilot, Generative AI by Adobe, Rytr Deep learning, and OpenAI API Key.

Table 2. Frequencies and percentages.

Category	Count	Percent
ChatGPT	274	74.7
QuillBot	135	36.8
Grammarly	248	67.6
Scholarcy	36	9.8
Scite	43	11.7
pdf.ai	68	18.5
Other	24	6.5

4.14.3. How Has Management Supported the Usage of AI in Your Workplace?

Table 3 shows that (21.3%) of all respondents agreed that management supported the usage of AI in their workplace by conferences, (29.4%) of all respondents agreed that management supported the usage of AI in their workplace by workshops, (34.9%) of all respondents agreed that management supported the usage of AI in their workplace by training, (22.6%) of all respondents agreed that management supported the usage of AI in their workplace by all of conferences, workshops, and training, and finally (18.8%) of all respondents agreed that management supported the usage of AI in their workplace in other ways.

Table 3. Frequencies and Percentages of Management Support.

Category	Count	Percent
Conferences	78	21.3
Workshops	108	29.4
Training	128	34.9
All of the above	83	22.6
Other	69	18.8

4.14.4. What Are Some of the Resources That You Believe Support the Adoption of AI in Your Organization?

Table 4 shows that (19.1%) of all respondents believe that the resources that support the adoption of AI in their organization are application processes, (15.3%) of all respondents believe that the resources support the adoption of AI in their organization are collaboration strategies, (23.2%) of all respondents believe that the resources support the adoption of AI in their organization are IT development plans, (26.4%) of all respondents believe that the resources support the adoption of AI in their organization are technical knowledge/skills, (48.2%) of all respondents believe that the resources support the adoption of AI in their organization, and finally, (2.5%) of all respondents believe that other resources are supporting the adoption of AI in their organization.

Table 4. Frequencies and Percentages Resources That Believeed to Support the Adoption.

Category	Count	Percent
Application processes	70	19.1
Collaboration strategies	56	15.3
IT development plans	85	23.2
technical knowledge/skills	97	26.4
All of the above	177	48.2
Other	9	2.5

4.14.5. What Are Some of the Assistances Offred by State Authorities to Motivate the Adoption of AI?

Table 5 shows that (26.4%) of all respondents believe that the social attitudes about morals and ethics offered by state authorities motivate the adoption of AI, (19.3%) of all respondents believe that the guidelines for the development of AI applications offered by state authorities motivate the adoption of AI, (33.5%) of all respondents believe that the protect privacy and ownership rights offered by state authorities motivates the adoption of AI, (28.1%) of all respondents believe that all of the above resources offered by state authorities motivate the adoption of AI, and finally (13.1%) of all respondents believe that other resources offered by state authorities motivate the adoption of AI.

Table 5. Frequencies and Percentages of Assisted Offenders by State Authorities.

Category	Count	Percent
Social attitudes about morals and ethical concerns	97	26.4
Offer guidelines for the development of AI applications	71	19.3
Protect Privacy and Ownership rights	123	33.5
All of the above	103	28.1
Other	48	13.1

4.14.6. What Technological Support Does Your Organization Have to Support the Adoption of AI?

Table 6 shows that (24.3%) of all respondents believe that their organization has supportive AI in-house software to support the adoption of AI, (20.2%) of all respondents believe that their organization has adoptive operating systems that support AI to support the adoption of AI, (20.4%) of all respondents believe that their organization has a supportive in-house AI network to support the adoption of AI, (47.7%) of all respondents believe that their organization is not yet there, none of the above support the adoption of AI, finally (2.2%) of all respondents believe that their organization has other technological support to support the adoption of AI.

Table 6. Frequencies and Percentages Technological Support.

Category	Count	Percent
Supportive AI in-house software.	89	24.3
Adoptive operating systems that support AI.	74	20.2
Supportive AI in-house Network.	75	20.4
Not yet there, none of the above.	175	47.7
Other	8	2.2

4.15. Testing the Model

4.15.1. Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) is used to validate the factor structure of the collection of observed variables (the factor loadings). Convergence validity and composite reliability (CR) are evaluated. Table 7 below displays the findings. Discriminate validity is seen in Table 8.

Given that the recommended factor loading is 0.50 or higher, and ideally 0.70 or higher [94], Table 7 demonstrates that all of the item loadings range from 0.621 to 0.874; the results are therefore accepted.

Composite reliability (CR) and average variance extracted (AVE) can be used to evaluate convergent validity in factor loadings. According to the findings, composite reliability scores of 0.757 to 0.905, which are higher than 0.7, indicate strong internal consistency. Additionally, the results demonstrate that the average variance extracted (AVE) values, which are greater than 0.50 (the cut-off value justifies the usage of the construct), ranged from 0.512 to 0.714. As a result, all of the latent variables satisfy the requirements needed to demonstrate convergent validity [15].

Table 7. Confirmatory factor analysis results (factor loading).

Latent Variable	Indicator	FL	FLS	AVE (>0.50)	CR (>0.70)	Cronbach's Alpha
Compatibility	C1	0.82	0.672	0.585	0.875	0.883
	C2	0.663	0.440			
	C3	0.831	0.691			
	C4	0.765	0.585			
	C5	0.732	0.536			
Complexity	CX1	0.873	0.762	0.574	0.843	0.867
	CX2	0.698	0.487			
	CX3	0.753	0.567			
	CX4	0.694	0.482			
User User Experience	UX1	0.867	0.752	0.697	0.902	0.938
	UX2	0.839	0.704			
	UX3	0.848	0.719			
	UX4	0.784	0.615			
Ease of Use	PEOU1	0.874	0.764	0.585	0.807	0.821
	PEOU2	0.721	0.520			
	PEOU3	0.687	0.472			
User Satisfaction	US1	0.763	0.582	0.615	0.905	0.95
	US2	0.721	0.520			
	US3	0.738	0.545			
	US4	0.865	0.748			
	US5	0.832	0.692			
	US6	0.778	0.605			
Performance Expectation	PE1	0.757	0.573	0.664	0.855	0.881
	PE2	0.811	0.658			
	PE3	0.872	0.760			

Table 7. *Cont.*

Latent Variable	Indicator	FL	FLS	AVE (>0.50)	CR (>0.70)	Cronbach's Alpha
AI Strategic Alignment	AIS1	0.834	0.696	0.573	0.80	0.835
	AIS2	0.757	0.573			
	AIS3	0.671	0.450			
Availability of Resources	AVR1	0.704	0.496	0.614	0.826	0.862
	AVR2	0.785	0.616			
	AVR3	0.854	0.729			
Competitive Pressure	COP1	0.716	0.513	0.555	0.789	0.817
	COP2	0.765	0.585			
	COP3	0.754	0.569			
Government Regulations	GOR1	0.784	0.615	0.528	0.77	0.814
	GOR2	0.682	0.465			
	GOR3	0.711	0.506			
Technological Support	TS1	0.621	0.386	0.512	0.757	0.805
	TS2	0.745	0.555			
	TS3	0.772	0.596			
Facilitating Conditions	FC1	0.857	0.734	0.709	0.88	0.913
	FC2	0.823	0.677			
	FC3	0.846	0.716			
AI Adoption Intentions	AIA1	0.844	0.712	0.714	0.882	0.929
	AIA2	0.856	0.733			
	AIA3	0.834	0.696			

FL = Factor Loading, FLS = Factor Loading Squared, AVE = Average Variance Extracted, CR = Composite Reliability. Refer to Appendix B for the full list of abbreviations.

Table 8. HTTP analysis.

	C	CX	UX	PEOU	US	PE	AINT	AIS	AVR	COP	GOR	TS	
C													
CX	0.722												
UX	0.815	0.834											
PEOU	0.772	0.785	0.822										
US	0.624	0.523	0.613	0.561									
PE	0.795	0.793	0.752	0.784	0.712								
AINT	0.661	0.699	0.771	0.693	0.734	0.653							
AIS	0.534	0.512	0.535	0.453	0.597	0.491	0.482						
AVR	0.514	0.543	0.533	0.618	0.499	0.487	0.667	0.559					
COP	0.732	0.702	0.688	0.725	0.694	0.596	0.738	0.644	0.723				
GOR	0.504	0.478	0.468	0.574	0.489	0.555	0.573	0.513	0.497	0.533			
TS	0.706	0.665	0.624	0.718	0.677	0.649	0.746	0.728	0.625	0.729	0.634		
FC	0.783	0.798	0.809	0.739	0.77	0.759	0.61	0.812	0.799	0.675	0.755	0.822	

All of the HTMT values obtained are less than 0.85, as shown in Table 8, suggesting that there are no issues with discriminant validity. Henseler et al. (2015) state that discriminant validity amongst reflective constructs is established by HTMT values less than 0.90 [95].

According to the findings, there were no overlapping items in the impacted constructs according to respondents' perceptions, and there were no collinearity issues among the latent constructs (multicollinearity).

Based on the results of Tables 7 and 8 above, the final best-fitting model is presented in Figure 2 below.

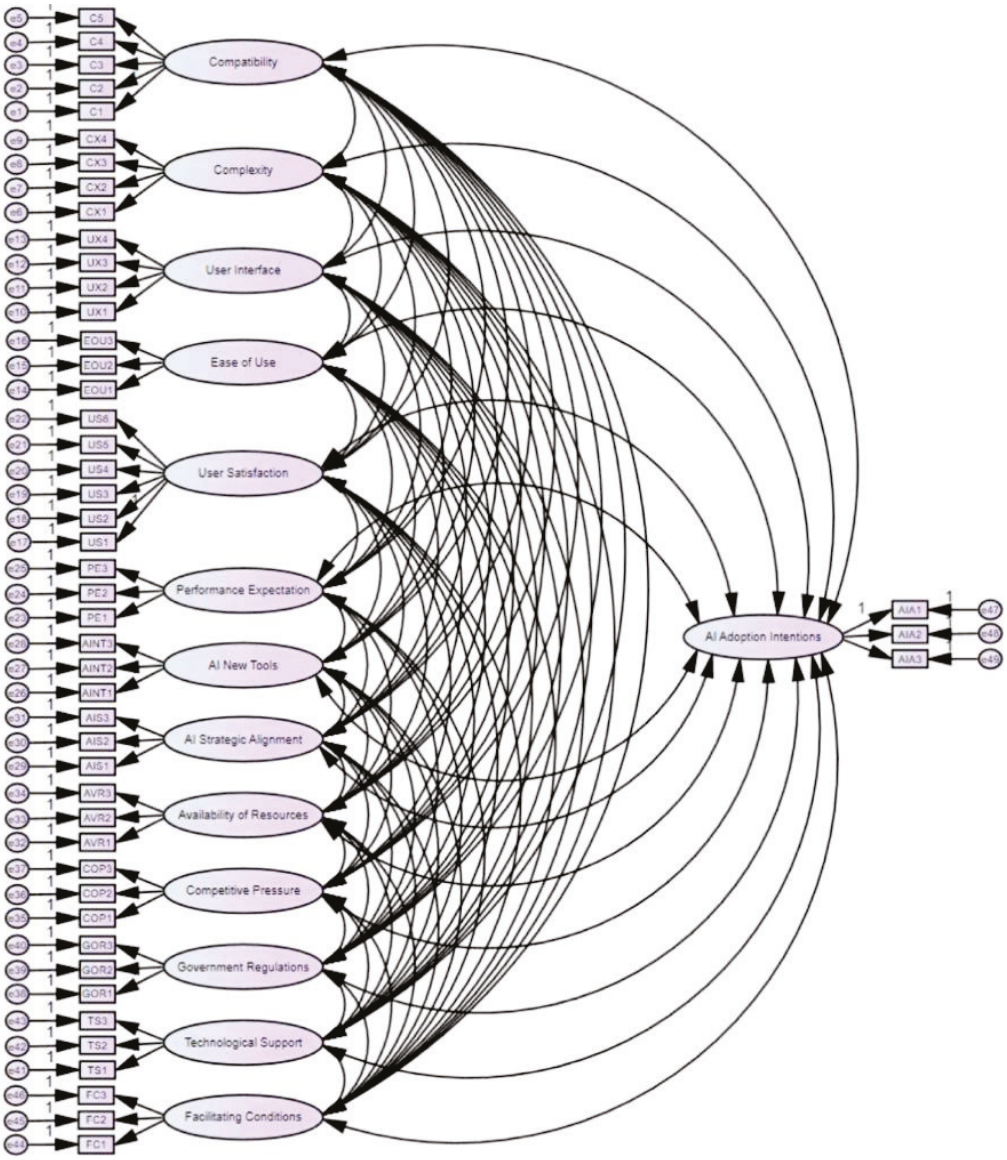


Figure 2. Final best-fitting CFA model.

4.15.2. Goodness of Fit

Several metrics are used to assess the model's goodness of fit, Standardized Root Mean Squared Residual (SRMR), comparative fit index (CFI), Tucker and Lewis's index of fit (TLI), normed fit index (NFI), and root mean square error of approximation (RMSEA)). Other indicators include the recommended cut-off values of model fit (Chi-square χ^2 ($p > 0.05$); Normed Chi-Square (χ^2/df) $1.0 \leq \chi^2/\text{df} \leq 3$; RMSEA ≤ 0.10, NFI ≥ 0.90; CFI ≥ 0.90; IFI ≥ 0.90; TLI ≥ 0.90). Table 9 below displays the findings.

Table 9. Final measurements of model fit.

X^2	51.213
X^2/DF	5.12
SRMR	0.037
CFI	0.951
TLI	0.924
NFI	0.958
IFI	0.958
RMSEA	0.07

Table 9 demonstrates that an excellent model fit is indicated by the SRMR value, which is less than 0.08 [18]. A great match for the model is shown by a CFI score greater than 0.95 [16]. Additionally, an excellent match is shown by the TLI value, which is greater than 0.90 [96]. A good match for the model is also shown by the NFI and IFI values, both of which are greater than 0.90 [18]. A good fit for the model is indicated when the RMSEA is less than or equal to 0.1 [97].

The suggested model is fitted since indexes indicate that it adequately fits the available data.

4.16. Testing the Hypotheses

A covariance-based Structural Equation Model (CB-SEM), using Partial Least Squares, is utilized to evaluate the research hypotheses which are necessary for this study. Figure 3 shows the SEM model hypotheses.

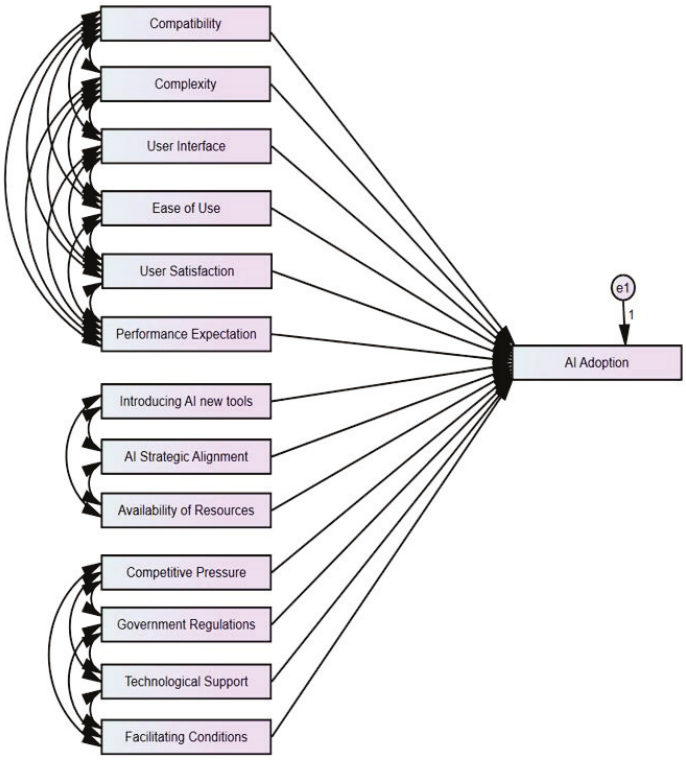


Figure 3. The CB-SEMmodel for the hypothesis.

4.17. Testing the First Hypothesis

H1: *There is a statistically significant impact of key factors ("C", "CX", "UX", "PEOU", "US", "PE", "AINT", "AIS", "AVR", "COP", "GOR", "TS", and "FC") together on AI adoption intentions. at a level of* $p \leq 0.05$.

This hypothesis was divided into 13 sub-hypotheses:

H1a: *There is a statistically significant impact of "C" on AI adoption intentions at a level of* $p \leq 0.05$.

H1b: *There is a statistically significant impact of "CX" on AI adoption intentions at a level of* $p \leq 0.05$.

H1c: *There is a statistically significant impact of User Experience "UX" on AI adoption intentions at a level of* $p \leq 0.05$.

H1d: *There is a statistically significant impact of "PEOU" on AI adoption intentions at a level of* $p \leq 0.05$.

H1e: *There is a statistically significant impact of "US" on AI adoption intentions at a level of* $p \leq 0.05$.

H1f: *There is a statistically significant impact of "PE" on AI adoption intentions at a level of* $p \leq 0.05$.

H1g: *There is a statistically significant impact of "AINT" on AI adoption intentions at a level of* $p \leq 0.05$.

H1h: *There is a statistically significant impact of "AIS" on AI adoption intentions at a level of* $p \leq 0.05$.

H1i: *There is a statistically significant impact of (AVR" on AI adoption intentions at a level of* $p \leq 0.05$.

H1j: *There is a statistically significant impact of (COP" on AI adoption intentions at a level of* $p \leq 0.05$.

H1k: *There is a statistically significant impact of (GOR" on AI adoption intentions at a level of* $p \leq 0.05$.

H1l: *There is a statistically significant impact of "TS" on AI adoption intentions at a level of* $p \leq 0.05$.

H1m: *There is a statistically significant impact of (FC" on AI adoption intentions at a level of* $p \leq 0.05$.

The result of the SEM for testing the hypotheses is presented in Table 10 below, which shows the following results:

- "C" has a positive significant impact on AI adoption intentions, as indicated by the regression weights; the route is significant since the *p*-value (***) is less than 0.001 and

the crucial ratio value is more than 2 [17]. Consequently, it is decided to embrace the first alternative sub-hypothesis;

- (CX) has a significant positive impact on AI adoption intentions, as indicated by the regression weights; the route is significant since the *p*-value (***) is less than 0.001 and the crucial ratio value is more than 2 [17]. Consequently, it is decided to embrace the second alternative sub-hypothesis;
- (CX) has a significant positive impact on AI adoption intentions, as indicated by the regression weights; the route is significant since the *p*-value (***) is less than 0.001 and the crucial ratio value is more than 2 [17]. Consequently, it is decided to embrace the second alternative sub-hypothesis;
- User Experience (UX) has a significant positive impact on AI adoption intentions, as indicated by the regression weights; the route is significant since the *p*-value (***) is less than 0.001 and the crucial ratio value is more than 2 [17]. Consequently, it is decided to embrace the third alternative sub-hypothesis;
- (PEOU) has a significant positive impact on AI adoption intentions, as indicated by the regression weights; the route is significant since the *p*-value (***) is less than 0.001 and the crucial ratio value is more than 2 [17]. Consequently, it is decided to embrace the fourth alternative sub-hypothesis;
- (US) has a significant positive impact on AI adoption intentions, as indicated by the regression weights; the route is significant since the *p*-value (***) is less than 0.001 and the crucial ratio value is more than 2 [17]. Consequently, it is decided to embrace the fifth alternative sub-hypothesis;
- (PE) has a positive significant impact on AI adoption intentions, as indicated by the regression weights; the route is significant since the *p*-value (0.001) is less than 0.01 and the crucial ratio value is more than 2 [17]. Consequently, it is decided to embrace the sixth alternative sub-hypothesis;
- (AINT) has a significant positive impact on AI adoption intentions, as indicated by the regression weights; the route is significant since the *p*-value (***) is less than 0.001 and the crucial ratio value is more than 2 [17]. Consequently, it is decided to embrace the seventh alternative sub-hypothesis;
- (AIS) has a significant positive impact on AI adoption intentions, as indicated by the regression weights; the route is significant since the *p*-value (0.003) is less than 0.01 and the crucial ratio value is more than 2 [17]. Consequently, it is decided to embrace the eighth alternative sub-hypothesis;
- (AVR) has a significant positive impact on AI adoption intentions, as indicated by the regression weights; the route is significant since the *p*-value (***) is less than 0.001 and the crucial ratio value is more than 2 [17]. Consequently, it is decided to embrace the ninth alternative sub-hypothesis;
- As per Byrne (2013), the regression weights indicate that (COP) has an insignificant impact on AI adoption intentions. This is because the critical ratio value is less than 2, and the *p*-value (0.421) is higher than 0.05, indicating that the path is not significant [17]. The tenth null hypothesis is thus accepted;
- As per Byrne (2013), the regression weights indicate that (GOR) has an insignificant impact on AI adoption intentions. This is because the critical ratio value is less than 2, and the *p*-value (0.785) is higher than 0.05, indicating that the path is not significant [17]. The eleventh null hypothesis is thus accepted;
- (TS) has a significant positive impact on AI adoption intentions, as indicated by the regression weights; the route is significant since the *p*-value (***) is less than 0.001 and the crucial ratio value is more than 2 [17]. Consequently, it is decided to embrace the twelfth alternative sub-hypothesis;

- (FC) has a significant positive impact on AI adoption intentions, as indicated by the regression weights; the route is significant since the *p*-value (***) is less than 0.001 and the crucial ratio value is more than 2 [17]. Consequently, it is decided to embrace the thirteenth alternative sub-hypothesis.

Table 10. Structural equation modelling regression weights.

				Estimate	S.E.	C.R.	*p*	Result
H1a	C	→	AIA	0.342	0.054	6.876	***	Not Supported
H1b	CX	→	AIA	0.268	0.044	6.085	***	Supported
H1c	UX	→	AIA	0.421	0.058	8.154	***	Not Supported
H1d	PEOU	→	AIA	0.332	0.045	7.382	***	Supported
H1e	US	→	AIA	0.216	0.046	4.672	***	Supported
H1f	PE	→	AIA	0.186	0.043	4.312	0.001	Supported
H1g	AINT	→	AIA	0.766	0.033	23.519	***	Supported
H1h	AIS	→	AIA	0.100	0.031	3.263	0.003	Supported
H1i	AVR	→	AIA	0.122	0.022	5.587	***	Supported
H1j	COP	→	AIA	0.072	0.035	1.004	0.421	Not Supported
H1k	GOR	→	AIA	0.008	0.029	0.743	0.785	Not Supported
H1l	TS	→	AIA	0.551	0.034	8.581	***	Not Supported
H1m	FC	→	AIA	0.964	0.039	25.000	***	Supported

S.E. = Standard errors of the regression weights, C.R. = Critical Ratio, *p* = *p*-value (*** < 0.001).

4.18. Testing the Second Hypothesis

H2: *Demographic factors (gender, age, education, major, and years of experience) moderate the original relationship between key factors "C", "CX", User Interface, "PEOU", "US", "PE", "AINT", "AIS", "AVR", "GOR", "TS", and "FC") together and AI adoption intentions.*

H2a: *Gender significantly moderates the relationship between key factors and AI adoption intentions at* $p \leq 0.05$.

H2b: *Age significantly moderates the relationship between key factors and AI adoption intentions at* $p \leq 0.05$

H2c: *Education level significantly moderates the relationship between key factors and AI adoption intentions at* $p \leq 0.05$.

H2d: *Academic major significantly moderates the relationship between key factors and AI adoption intentions at* $p \leq 0.05$.

H2e: *Years of experience significantly moderates the relationship between key factors and AI adoption intentions at* $p \leq 0.05$.

The second main hypothesis is tested through multiple-group CB-SEM analysis using AMOS for the seven demographics it represents.

The results of the sub-hypothesis testing are presented in the following subsections.

H2a: *Gender Moderation: Gender categorical moderation is examined, and the results are presented in Table 11 below.*

Table 11. Multiple-group SEM analysis results for gender model.

Model	Structural Weights
DF	1
CMIN	0.455
p	0.491
NFI Delta-1	0.002
IFI Delta-2	0.002

Table 11 demonstrates that because the *p*-value (0.491) is higher than (0.05), the chi-square value (0.455) is not significant. This indicates that the disparities between the groups of men and women are negligible.

H2b: *Age Moderation: Age categorical moderation is examined, and the results are presented in Table 12 below.*

Table 12. Multiple-group SEM analysis results for age model.

Model	Structural Weights
DF	2
CMIN	1.279
p	0.322
NFI Delta-1	0.009
IFI Delta-2	0.009

Given that the *p*-value (0.322) is higher than (0.05), Table 12 demonstrates that the chi-square value (1.279) is not significant. This suggests that age has no discernible moderating influence on the first association between the AI key factors ("C", "CX", "UX", "PEOU", "US", "PE", "AINT", "AIS", "AVR", "GOR", "TS", and "FC") together and AI adoption intentions because there are no notable variations across the various age groups.

H2c: *Education Moderation: The Education categorical moderation is examined, and the results are presented in Table 13 below.*

Table 13. Multiple-group SEM analysis results for education model.

Model	Structural Weights
DF	2
CMIN	4.624
p	0.099
NFI Delta-1	0.017
IFI Delta-2	0.017

Given that the *p*-value (0.099) is higher than (0.05), Tables 4–13 demonstrates that the chi-square value (4.624) is not significant. This suggests that education has no discernible moderating influence on the first association between the AI key factors ("C", "CX", "UX", "PEOU", "US", "PE", "AINT", "AIS", "AVR", "COP", "GOR", "TS", and "FC") together and AI adoption intentions, because there are no notable variations across the various education level groups.

H2d: *Major Moderation: The Major categorical moderation is examined, and the results are presented in Table 14 below.*

Table 14. Multiple-group SEM analysis results for major model.

Model	Structural Weights
DF	4
CMIN	12.939
p	0.012
NFI Delta-1	0.053
IFI Delta-2	0.053

Table 14 shows that the chi-square value (12.939) is significant since the *p*-value (0.012) is less than (0.05). This means significant differences exist between the different types of major groups. This suggests that major has a moderating influence on the first association between the AI key factors ("C", "CX", "UX", "PEOU", "US", "PE", "AINT", "AIS", "AVR", "COP", "GOR", "TS", and "FC") together and AI adoption intentions.

While the results in Table 15 show that all types of majors have a significant moderation effect since the critical ratio value is greater than 2 and the *p*-values are less than 0.01, the path is significant, except for the medicine major which has an insignificant moderation effect on the original relationships between the AI key factors and AI adoption intentions.

Table 15. Structural Equation Modelling Regression Weights of AI Key Factors.

			Estimate	S.E.	C.R.	*p*	Effect	R^2
AI Key Factors (Less than 2 years)	→	AIA	0.606	0.222	2.726	0.006	0.339	0.115
AI Key Factors (2 years–less than 6 years)	→	AIA	1.136	0.145	7.839	***	0.666	0.444
AI Key Factors (6 years–less than 8 years)	→	AIA	1.481	0.273	5.429	***	0.738	0.544
AI Key Factors (8 years–less than 10 years)	→	AIA	1.366	0.098	13.886	***	0.907	0.823
AI Key Factors (10 years and above)	→	AIA	1.195	0.080	14.890	***	0.760	0.578

S.E. = Standard errors of the regression weights, C.R. = Critical Ratio, *p* = *p*-value (*** < 0.001).

Other majors have the biggest effect with 0.764, then come, respectively, (IT, Management, and Pharmaceutical) with effect values of (0.692, 0.676, and 0.675).

H2e: *Experience Moderation: The Experience categorical moderation is examined, and the results are presented in Table 16 below.*

Table 16. Multiple-group SEM analysis results for experience model.

			Estimate	S.E.	C.R.	*p*	Effect	R^2
AI Key Factors (IT)	→	AIA	1.053	0.095	11.110	***	0.692	0.479
AI Key Factors (Management)	→	AIA	1.576	0.201	7.829	***	0.676	0.456
AI Key Factors (Medicine)	→	AIA	0.219	0.338	0.648	0.517	0.138	0.019
AI Key Factors (Pharmaceutical)	→	AIA	1.275	0.422	3.017	0.003	0.675	0.456
AI Key Factors (Other)	→	AIA	1.250	0.097	12.890	***	0.764	0.584

S.E. = Standard errors of the regression weights, C.R. = Critical Ratio, *p* = *p*-value (*** < 0.001).

Table 16 shows that the chi-square value (10.625) is significant since the *p*-value (0.031) is less than (0.05). This means that there are significant differences between the different groups of Experience; therefore, the groups of Experience have a moderating influence on the first association between the AI key factors ("C", "CX", "UX", "PEOU", "US", "PE", "AINT", "AIS", "AVR", "COP", "GOR", "TS", and "FC") together and AI adoption intentions.

The results above in Table 17 show that all categories included in years of experience have a significant moderation effect since the critical ratio value is greater than 2 and the *p*-values are less than 0.01, the path is significant.

Table 17. Structural equation modelling regression weights.

Model	Structural Weights
DF	4
CMIN	10.625
p	0.03
NFI Delta-1	0.038
IFI Delta-2	0.038

Years of experience (8 years–less than 10 years) has the biggest effect, with 0.907, then come years of experience (10 years and above, 6 years–less than 8 years, 2 years–less than 6 years, and less than 2 years) with effect values (0.760, 0.738, 0.666, and 0.339).

5. Conclusions and Future Work

5.1. Conclusions

The findings of this study support several factors that trigger higher education institutions to innovate and adopt Artificial Intelligence. The study revealed that all investigated factors, namely "C", "CX", "UX", "PEOU", "US", "PE", AI introduces new tools "AINT", "AIS", "AVR", "TS", and "FC", have a statistically significant positive impact on AI adoption intentions. The statistical analysis showed that the *p*-values for all factors reached high significance ($p < 0.001$ for most factors), allowing us to accept the alternative hypothesis for the aforementioned factors. However, "COP" and "GOR" showed no statistically signifi-

cant impact on AI adoption intentions, leading us to accept the null hypothesis for those factors. Demographic factors were also examined as possible moderators in this study. It was found that age and education level did not significantly impact the relationship between key factors and AI adoption intentions. Significant moderation effects were observed for main fields of study and years of experience, suggesting that these demographic factors might influence how individuals perceive and formulate their intentions to adopt AI in a higher education context.

5.2. Future Work and Recommendations

The results of this study offer a useful understanding of the elements that influence the intent of higher education institutions to adopt AI technology. Considering the notable effects that have been identified, several suggestions can be put forth for future endeavors and real-world implementations:

(1) Compatibility "C": The results indicate that compatibility has a significant positive impact on AI adoption intentions. Further research should investigate how institutions might improve the compatibility of AI technology with current systems and processes to allow a more effortless adoption.
(2) "CX": complexity also shows a significant positive impact on AI adoption intentions. Further study endeavors may explore methods to streamline AI technologies and diminish apparent intricacy, promoting wider consumer acceptance.
(3) "UX": The positive impact of user interface on AI adoption aspirations underscores the need to craft user-friendly interfaces. Subsequent research should prioritize creating user-friendly and easily available artificial intelligence systems that address the varied requirements of individuals in higher education.
(4) "PEOU": The strong correlation between "PEOU" and AI adoption intentions indicates that institutions should prioritize providing training and support to boost users' confidence in employing AI technologies. Subsequent studies could investigate the efficacy of various training programs in enhancing the perception of usability.
(5) "US": User satisfaction significantly influences AI adoption intentions, indicating that organizations must ensure a positive user experience with AI tools. Subsequent research should investigate the elements influencing user satisfaction and determine methods.
(6) "PE": The findings reveal that "PE" positively impacts AI adoption intentions. Future research should explore how organizations might effectively convey the anticipated advantages of AI technologies to prospective users.
(7) Demographic variables: The study highlights the mediating roles of demographic variables such as age, gender, education, and years of experience. Further investigation is needed to explore the impact of these characteristics on the adoption of AI technology and develop strategies accordingly. To summarize, the results of this study highlight the significance of resolving the highlighted elements to improve the intent of higher education institutions to use artificial intelligence. Further investigation should be conducted to examine these aspects, offering practical knowledge for policymakers and educational administrators to promote the effective incorporation of AI in academic environments.

5.3. Practical Implications

The research findings offer higher education institutions looking to adopt AI-enabled technologies several useful takeaways. First, by integrating AI tools with current teaching and learning methods, university administrators should prioritize improving the perceived utility and compatibility of these tools. Adoption rates can be greatly increased by training

programs that increase faculty and staff confidence in utilizing AI technologies, especially by addressing issues with usability and complexity. Institutions should also make investments in strong technology infrastructure and guarantee that sufficient resources and support systems are available. Decision-makers at the policy level must take demographic variations into account and create inclusive strategies that encourage adoption by people with a range of educational backgrounds and levels of experience. These doable actions can hasten the effective adoption of AI in settings related to higher education.

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

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to restrictions related to participant privacy and ethical considerations.

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

Appendix A

Table A1. Survey Items and Sources.

Construct	Item Codes	Source
Compatibility	C1–C5	Moore and Benbasat (1991) [98]
Complexity	CX1–CX4	Rogers (2003) [11]
User Experience (UX)	UX1–UX4	Zhang and Adipat (2005) [99]
Perceived Ease of Use	PEOU1–PEOU3	Davis (1989) [14]
User Satisfaction	US1–US6	DeLone and McLean (2003) [100]
Performance Expectation	PE1–PE3	Venkatesh et al. (2003) [82]
AI Strategic Alignment	AIS1–AIS3	Luftman et al. (2004) [101]
Availability of Resources	AVR1–AVR3	Tornatzky and Fleischer (1990) [12]
Competitive Pressure	COP1–COP3	Zhu and Kraemer (2005) [102]
Government Regulations	GOR1–GOR3	Kuan and Chau (2001) [103]
Technological Support	TS1–TS3	Ifinedo (2011) [83]
Facilitating Conditions	FC1–FC3	Venkatesh et al. (2003) [82]
AI Adoption Intentions	AIA1–AIA3	Venkatesh and Bala (2008) [104]

Appendix B

Table A2. List of Abbreviations.

Abbreviation	Definition
AI	Artificial Intelligence
TAM	Technology Acceptance Model
DOI	Diffusion of Innovation
TOE	Technology–Organization–Environment
CB-SEM	Covariance-Based Structural Equation Modeling
PEOU	Perceived Ease of Use
UX	User Experience
US	User Satisfaction
PE	Performance Expectation
AIS	AI Strategic Alignment
AVR	Availability of Resources
COP	Competitive Pressure
GOR	Government Regulations
TS	Technological Support
FC	Facilitating Conditions
AIA	AI Adoption Intentions
AVE	Average Variance Extracted
CR	Composite Reliability
PU	Perceived usefulness
MLE	Maximum Likelihood Estimation
PLS-SEM	Partial Least Squares Structural Equation Modeling.

References

1. Volberda, H.W.; Khanagha, S.; Baden-Fuller, C.; Mihalache, O.R.; Birkinshaw, J. Strategizing in a digital world: Overcoming cognitive barriers, reconfiguring routines and introducing new organizational forms. *Long Range Plan.* **2021**, *54*, 102110. [CrossRef]
2. Duan, Y.; Edwards, J.S.; Dwivedi, Y.K. Artificial intelligence for decision making in the era of Big Data-evolution, challenges and research agenda. *Int. J. Inf. Manag.* **2019**, *48*, 63–71. [CrossRef]
3. Jöhnk, J. Managing digital transformation: Challenges and choices in organizational design and decision-making. Ph.D. Thesis, University of Bayreuth, Bayreuth, Germany,, 2020. [CrossRef]
4. Pillai, R.; Sivathanu, B. Adoption of artificial intelligence (ai) for talent acquisition in it/ites organizations. *Benchmarking Int. J.* **2020**, *27*, 2599–2629. [CrossRef]
5. Chatterjee, S.; Ghosh, S.K.; Chaudhuri, R. Knowledge management in improving business process: An interpretative framework for successful implementation of ai-crm-km system in organizations. *Bus. Process Manag. J.* **2020**, *26*, 1261–1281. [CrossRef]
6. Bharadiya, J.P. Machine learning and AI in business intelligence: Trends and opportunities. *Int. J. Comput. (IJC)* **2023**, *48*, 123–134.
7. Enholm, I.M.; Papagiannidis, E.; Mikalef, P.; Krogstie, J. Artificial intelligence and business value: A literature review. *Inf. Syst. Front.* **2022**, *24*, 1709–1734. [CrossRef]
8. Dwivedi, Y.K.; Hughes, L.; Ismagilova, E.; Aarts, G.; Coombs, C.; Crick, T.; Williams, D.M. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int. J. Inf. Manag.* **2021**, *57*, 101994. [CrossRef]
9. Chatterjee, S.; Ghosh, S.K.; Chaudhuri, R.; Chaudhuri, S. Adoption of ai integrated crm system by indian industry: From security and privacy perspective. *Comput. Secur.* **2020**, *29*, 1–24. [CrossRef]
10. George, B.; Wooden, O. Managing the strategic transformation of higher education through artificial intelligence. *Adm. Sci.* **2023**, *13*, 196. [CrossRef]

11. Rogers, E.M. *Diffusion of Innovations*; Free Press: New York, NY, USA, 2003.
12. Tornatzky, L.G.; Fleischer, M. *The Processes of Technological Innovation*; Issues in Organization and Management Series; Lexington Books: Lexington, MA, USA, 1990.
13. Venkatesh, V.; Davis, F.D. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Manag. Sci.* **2000**, *46*, 186–204. [CrossRef]
14. Davis, F.D.; Bagozzi, R.P.; Warshaw, P.R. Technology acceptance model. *J. Manag. Sci.* **1989**, *35*, 982–1003.
15. Hair, J.F.; Ringle, C.M.; Sarstedt, M. CB-SEM: Indeed a silver bullet. *J. Mark. Theory Pract.* **2011**, *19*, 139–152. [CrossRef]
16. Kline, R.B. *Principles and Practice of Structural Equation Modeling*; Guilford Press: New York, NY, USA, 2005.
17. Byrne, B.M. Adaptation of assessment scales in cross-national research: Issues, guidelines, and caveats. *Int. Perspect. Psychol.* **2016**, *5*, 51–65. [CrossRef]
18. Hu, L.T.; Bentler, P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. Multidiscip. J.* **1999**, *6*, 1–55. [CrossRef]
19. Greenhalgh, T.; Robert, G.; Macfarlane, F.; Bate, P.; Kyriakidou, O. Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Q.* **2004**, *82*, 581–629. [CrossRef]
20. Bozkurt, A.; Karadeniz, A.; Bañeres, D.; Rodríguez, M.E. Artificial intelligence and reflections from educational landscape: A review of ai studies in half a century. *Sustainability* **2021**, *13*, 800. [CrossRef]
21. Crompton, H.; Song, D. El potencial de la inteligencia artificial en la educación superior. *Rev. Virtual Univ. Católica Norte* **2021**, *62*, 1–4. [CrossRef]
22. Jain, R.; Garg, N.; Khera, S.N. Adoption of ai-enabled tools in social development organizations in india: An extension of utaut model. *Front. Psychol.* **2022**, *13*, 893691. [CrossRef]
23. Popenici, S.; Kerr, S. Exploring the impact of artificial intelligence on teaching and learning in higher education. *Technol. Enhanc. Learn.* **2017**, *12*, 22. [CrossRef]
24. Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. Systematic review of research on artificial intelligence applications in higher education-where are the educators? *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 39. [CrossRef]
25. Wu, W.; Zhang, B.; Li, S.; Liu, H. Exploring Factors of the Willingness to Accept AI-Assisted Learning Environments: An Empirical Investigation Based on the UTAUT Model and Perceived Risk Theory. *Front. Psychol.* **2022**, *13*, 870777. [CrossRef] [PubMed]
26. Hannan, E. Ai: New source of competitiveness in higher education. *Compet. Rev. Int. Bus. J.* **2021**, *33*, 265–279. [CrossRef]
27. Chen, L.; Chen, P.; Lin, Z. Artificial intelligence in education: A review. *IEEE Access* **2020**, *8*, 75264–75278. [CrossRef]
28. Luckin, R.; Cukurova, M. Designing educational technologies in the age of ai: A learning sciences-driven approach. *Br. J. Educ. Technol.* **2019**, *50*, 2824–2838. [CrossRef]
29. Low, C.; Chen, Y.; Wu, M. Understanding the determinants of cloud computing adoption. *Ind. Manag. Data Syst.* **2011**, *111*, 1006–1023.
30. Arshad, Y.; Chin, W.P.; Yahaya, S.N.; Nizam, N.Z.; Masrom, N.R.; Ibrahim, S.N.S. Small and medium enterprises' adoption for e-commerce in Malaysia tourism state. *Int. J. Acad. Res. Bus. Soc. Sci.* **2018**, *8*, 1457–1557. [CrossRef]
31. Pillai, R.; Metri, B.A.; Kaushik, N. Students' adoption of ai-based teacherbots (t-bots) for learning in higher education. *Inf. Technol. Amp. People* **2023**, *37*, 328–355. [CrossRef]
32. Paton, C.; Kobayashi, S. An open science approach to artificial intelligence in healthcare. *Yearb. Med. Inform.* **2019**, *28*, 47–051. [CrossRef]
33. Tuffaha, M.; Perello-Marin, M.R. Adoption factors of artificial intelligence in human resources management. *Future Bus. Adm.* **2022**, *1*, 1–12. [CrossRef]
34. AL-Takhayneh, S.K.; Karaki, W.; Hasan, R.A.; Chang, B.; Shaikh, J.M.; Kanwal, W. Teachers' psychological resistance to digital innovation in Jordanian entrepreneurship and business schools: Moderation of teachers' psychology and attitude toward educational technologies. *Front. Psychol.* **2022**, *13*, 1004078. [CrossRef]
35. Islam, M.N.; Khan, N.I.; Inan, T.T.; Sarker, I.H. Designing user interfaces for illiterate and semi-literate users: A systematic review and future research agenda. *SAGE Open* **2023**, *13*, 21582440231172741. [CrossRef]
36. Ismatullaev, U.V.U.; Kim, S.H. Review of the factors affecting acceptance of ai-infused systems. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2022**, *66*, 126–144. [CrossRef] [PubMed]
37. Lee, J.C.; Chen, X. Exploring users' adoption intentions in the evolution of artificial intelligence mobile banking applications: The intelligent and anthropomorphic perspectives. *Int. J. Bank Mark.* **2022**, *40*, 631–658. [CrossRef]
38. Amin, M.; Rezaei, S.; Abolghasemi, M. User satisfaction with mobile websites: The impact of perceived usefulness (PU), perceived ease of use (PEOU) and trust. *Nankai Bus. Rev. Int.* **2014**, *5*, 258–274. [CrossRef]
39. Chahal, J.; Rani, N. Exploring the acceptance for e-learning among higher education students in India: Combining technology acceptance model with external variables. *J. Comput. High. Educ.* **2022**, *34*, 844–867. [CrossRef]
40. Kuo, Y.C.; Walker, A.E.; Schroder, K.E.; Belland, B.R. Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. *Internet High. Educ.* **2014**, *20*, 35–50. [CrossRef]

41. Zhu, Y.; Wang, R.; Pu, C. "I am chatbot, your virtual mental health adviser." What drives citizens' satisfaction and continuance intention toward mental health chatbots during the covid-19 pandemic? An empirical study in china. *Digital Health* **2022**, *8*, 20552076221090031. [CrossRef]
42. Dora, M.; Kumar, A.; Mangla, S.K.; Pant, A.; Kamal, M.M. Critical success factors influencing artificial intelligence adoption in food supply chains. *Int. J. Prod. Res.* **2021**, *60*, 4621–4640. [CrossRef]
43. Sun, H.; Fang, Y.; Zou, H. Choosing a fit technology: Understanding mindfulness in technology adoption and continuance. *J. Assoc. Inf. Syst.* **2016**, *17*, 377–412. [CrossRef]
44. Alalwan, A.A.; Dwivedi, Y.K.; Rana, N.P. Factors influencing adoption of mobile banking by Jordanian bank customers: Extending utaut2 with trust. *Int. J. Inf. Manag.* **2017**, *37*, 99–110. [CrossRef]
45. Rasheed, H.M.W.; Yuanqiong, H.; Khizar, H.M.U.; Khalid, J. What drives the adoption of artificial intelligence among consumers in the hospitality sector: A systematic literature review and future agenda. *J. Hosp. Tour. Technol.* **2024**, *15*, 211–231. [CrossRef]
46. Tarhini, A.; Masa'deh, R.; Al-Busaidi, K.A.; Mohammed, A.B.; Maqableh, M. Factors influencing students' adoption of e-learning: A structural equation modeling approach. *J. Int. Educ. Bus.* **2017**, *10*, 164–182. [CrossRef]
47. Lee, H.; Lee, S.; Shin, J. An analysis on the satisfaction and perception of performance outcomes of the university information disclosure system. *Asia-Pac. J. Educ. Manag. Res.* **2020**, *5*, 49–56. [CrossRef]
48. Henke, J. Navigating the ai era: University communication strategies and perspectives on generative ai tools. *J. Sci. Commun.* **2024**, *23*, A05. [CrossRef]
49. Okunlaya, R.O.; Abdullah, N.S.; Alias, R.A. Artificial intelligence (ai) library services innovative conceptual framework for the digital transformation of university education. *Libr. Hi Tech* **2022**, *40*, 1869–1892. [CrossRef]
50. Gupta, V.; Gupta, C. Synchronizing innovation: Unveiling the synergy of need-based and curiosity-based experimentation in ai technology adoption for libraries. *Libr. Hi Tech News* **2023**, *40*, 15–17. [CrossRef]
51. Sallu, S.; Raehang, R.; Qammaddin, Q. Exploration of artificial intelligence (ai) application in higher education. *Archit. High Perform. Comput.* **2024**, *6*, 315–327. [CrossRef]
52. Saidakhror, G. The impact of artificial intelligence on higher education and the economics of information technology. *Int. J. Law Policy* **2024**, *2*, 1–6. [CrossRef]
53. Jarrahi, M.H.; Kenyon, S.; Brown, A.; Donahue, C.; Wicher, C. Artificial intelligence: A strategy to harness its power through organizational learning. *J. Bus. Strategy* **2022**, *44*, 126–135. [CrossRef]
54. Bearman, M.; Ajjawi, R. Learning to work with the black box: Pedagogy for a world with artificial intelligence. *Br. J. Educ. Technol.* **2023**, *54*, 1160–1173. [CrossRef]
55. Greiner, C.; Peisl, T.C.; Höpfl, F.; Beese, O. Acceptance of ai in semi-structured decision-making situations applying the four-sides model of communication-an empirical analysis focused on higher education. *Educ. Sci.* **2023**, *13*, 865. [CrossRef]
56. Boonsiritomachai, W.; Mcgrath, G.M.; Burgess, S. Exploring business intelligence and its depth of maturity in Thai SMEs. *Cogent Bus. Manag.* **2016**, *3*, 1220663. [CrossRef]
57. Hungund, S.; Mani, V. Benchmarking of factors influencing adoption of innovation in software product SMEs: An empirical evidence from India. *Benchmarking Int. J.* **2019**, *26*, 1451–1468. [CrossRef]
58. Alsheibani, S.; Messom, C.; Cheung, Y. Re-Thinking the Competitive Landscape of Artificial Intelligence. In Proceedings of the Hawaii International Conference on System Sciences (HICSS), Maui, HI, USA, 7–10 January 2020; pp. 5861–5870.
59. Porter, M.E.; Millar, V.E. *How Information Gives You Competitive Advantage*; Routledge: London, UK, 1985.
60. Alghamdi, M.I. Assessing factors affecting intention to adopt AI and ML: The case of the Jordanian retail industry. *Period. Eng. Nat. Sci. (PEN)* **2020**, *8*, 2516–2524. [CrossRef]
61. Pan, Y.; Froese, F.; Liu, N. The adoption of artificial intelligence in employee recruitment: The influence of contextual factors. *Int. J. Hum. Resour. Manag.* **2022**, *33*, 1125–1147. [CrossRef]
62. Wong, J.W.; Yap, K.H.A. Factors influencing the adoption of artificial intelligence in accounting among micro, small medium enterprises (msmes). *Quantum J. Soc. Sci. Humanit.* **2024**, *5*, 16–28. [CrossRef]
63. Ghani, E.K.; Ariffin, N.; Sukmadilaga, C. Factors influencing artificial intelligence adoption in publicly listed manufacturing companies: A technology, organisation, and environment approach. *Int. J. Appl. Econ.* **2022**, *14*, 108–117. [CrossRef]
64. Alexander, C.S.; Yarborough, M.; Smith, A. Who is responsible for 'responsible AI'?: Navigating challenges to build trust in AI agriculture and food system technology. *Precis. Agric.* **2024**, *25*, 146–185. [CrossRef]
65. Moon, M.J. Searching for inclusive artificial intelligence for social good: Participatory governance and policy recommendations for making ai more inclusive and benign for society. *Public Adm. Rev.* **2023**, *83*, 1496–1505. [CrossRef]
66. Farida, I.; Ningsih, W.; Lutfiani, N.; Aini, Q.; Harahap, E.P. Responsible urban innovation working with local authorities a framework for artificial intelligence (ai). *Sci. J. Inform.* **2023**, *10*, 121–126. [CrossRef]
67. Noordt, C.V.; Misuraca, G. Exploratory insights on artificial intelligence for government in Europe. *Soc. Sci. Comput. Rev.* **2020**, *40*, 426–444. [CrossRef]

68. Mohsin, F.H.; Isa, N.M.; Ishak, K.; Salleh, H. Navigating the adoption of artificial intelligence in higher education. *Int. J. Bus. Technopreneurship (IJBT)* **2024**, *14*, 109–120. [CrossRef]
69. Bai, X. The role and challenges of artificial intelligence in information technology education. *Pac. Int. J.* **2024**, *7*, 86–92. [CrossRef]
70. Opesemowo, O.A.G.; Adekomaya, V. Harnessing artificial intelligence for advancing sustainable development goals in south africa's higher education system: A qualitative study. *Int. J. Learn. Teach. Educ. Res.* **2024**, *23*, 67–86. [CrossRef]
71. Polyportis, A. A longitudinal study on artificial intelligence adoption: Understanding the drivers of chatgpt usage behavior change in higher education. *Front. Artif. Intell.* **2024**, *6*, 1324398. [CrossRef]
72. Eftimov, L.; Kitanovikj, B. Unlocking the path to ai adoption: Antecedents to behavioral intentions in utilizing ai for effective job (re)design. *J. Hum. Resour. Manag.-HR Adv. Dev.* **2023**, *2023*, 123–134. [CrossRef]
73. Tanantong, T.; Wongras, P. A utaut-based framework for analyzing users' intention to adopt artificial intelligence in human resource recruitment: A case study of thailand. *Systems* **2024**, *12*, 28. [CrossRef]
74. Morrison, K. Artificial intelligence and the nhs: A qualitative exploration of the factors influencing adoption. *Future Healthc. J.* **2021**, *8*, 648–654. [CrossRef]
75. Chen, H.; Li, L.; Chen, Y. Explore success factors that impact artificial intelligence adoption on telecom industry in china. *J. Manag. Anal.* **2020**, *8*, 36–68. [CrossRef]
76. Chen, C.; Chen, S.; Khan, A.; Lim, M.K.; Tseng, M. Big data analytics-artificial intelligence and supply chain ambidexterity impacts on corporate image and green communication. *Ind. Manag. Data Syst.* **2024**, *124*, 2899–2918. [CrossRef]
77. Bughin, J. Does artificial intelligence kill employment growth: The missing link of corporate ai posture. *Front. Artif. Intell.* **2023**, *6*, 1239466. [CrossRef] [PubMed]
78. Govindan, K. How artificial intelligence drives sustainable frugal innovation: A multitheoretical perspective. *IEEE Trans. Eng. Manag.* **2024**, *71*, 638–655. [CrossRef]
79. Horani, O.M.; Al-Adwan, A.S.; Yaseen, H.; Hmoud, H.; Al-Rahmi, W.M.; Alkhalifah, A. The critical determinants impacting artificial intelligence adoption at the organizational level. *Inf. Dev.* **2023**, 02666669231166889. [CrossRef]
80. Faul, F.; Erdfelder, E.; Buchner, A.; Lang, A.-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods* **2009**, *41*, 1149–1160. [CrossRef]
81. Wut, T.M.; Chan, E.A.H.; Wong, H.S.M.; Chan, J.K. Perceived artificial intelligence literacy and employability of university students. *Educ. + Train.* **2025**, *67*, 258–274. [CrossRef]
82. Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User acceptance of information technology: Toward a unified view. *MIS Q.* **2003**, *27*, 425–478. [CrossRef]
83. Ifinedo, P. Internet/e-business technologies acceptance in Canada's SMEs: An exploratory investigation. *Internet Res.* **2011**, *21*, 255–281. [CrossRef]
84. Al-Gahtani, S.S. Modeling the electronic transactions acceptance using an extended technology acceptance model. *Appl. Comput. Inform.* **2011**, *9*, 47–77. [CrossRef]
85. Camisón, C.; Villar-López, A. Non-technical innovation: Organizational memory and learning capabilities as antecedent factors with effects on sustained competitive advantage. *Ind. Mark. Manag.* **2011**, *40*, 1294–1304. [CrossRef]
86. Harwood, S.; Eaves, S. Conceptualising technology, its development and future: The six genres of technology. *Technol. Forecast. Soc. Change* **2020**, *160*, 120174. [CrossRef]
87. Chang, H.C. A new perspective on Twitter hashtag use: Diffusion of innovation theory. *Proc. Am. Soc. Inf. Sci. Technol.* **2010**, *47*, 1–4. [CrossRef]
88. Ciftci, S.K.; Gok, R.; Karadag, E. Acceptance and use of the distance education systems of Turkish medical educators during COVID-19 pandemic: An analysis of contextual factors with the UTAUT2. *BMC Med. Educ.* **2023**, *23*, 36. [CrossRef] [PubMed]
89. Priyadarshinee, P.; Raut, R.D.; Jha, M.K.; Gardas, B.B. Understanding and predicting the determinants of cloud computing adoption: A two staged hybrid SEM-Neural networks approach. *Comput. Hum. Behav.* **2017**, *76*, 341–362. [CrossRef]
90. Park, Y.J.; Jeong, Y.J.; An, Y.S.; Ahn, J.K. Analyzing the Factors Influencing the Intention to Adopt Autonomous Ships Using the TOE Framework and DOI Theory. *J. Navig. Port Res.* **2022**, *46*, 134–144.
91. Ahmad, S.; Miskon, S.; Alkanhal, T.A.; Tlili, I. Modeling of business intelligence systems using the potential determinants and theories with the lens of individual, technological, organizational, and environmental contexts—A systematic literature review. *Appl. Sci.* **2020**, *10*, 3208. [CrossRef]
92. Beshdeleh, M.; Angel, A.; Bolour, L. Adoption of EBET Agency's Cloud Casino Software by Using TOE and DOI Theory as a Solution for Gambling Website. *J. Innov. Bus. Res.* **2020**, *116*, 100–119.
93. Qasem, Y.A.; Abdullah, R.; Yah, Y.; Atan, R.; Al-Sharafi, M.A.; Al-Emran, M. Towards the development of a comprehensive theoretical model for examining the cloud computing adoption at the organizational level. *Recent Adv. Intell. Syst. Smart Appl.* **2021**, *295*, 63–74.
94. Bollen, K.A. *Structural Equations with Latent Variables*; John Wiley & Sons: Hoboken, NJ, USA, 2014.

95. Henseler, J.; Ringle, C.M.; Sarstedt, M. A New Criterion for Assessing Discriminant Validity in Variance-based Structural Equation Modeling. *J. Acad. Mark. Sci.* **2015**, *43*, 115–135. [CrossRef]
96. Sharma, S.; Mukherjee, S.; Kumar, A.; Dillon, W.R. A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *J. Bus. Res.* **2005**, *58*, 935–943. [CrossRef]
97. Brown, T.A. *Confirmatory Factor Analysis for Applied Research*; Guilford Publications: New York, NY, USA, 2015.
98. Moore, G.C.; Benbasat, I. Development of an instrument to measure the perceptions of adopting an information technology innovation. *Inf. Syst. Res.* **1991**, *2*, 192–222. [CrossRef]
99. Zhang, D.; Adipat, B. Challenges, methodologies, and issues in the usability testing of mobile applications. *Int. J. Hum.-Comput. Interact.* **2005**, *18*, 293–308. [CrossRef]
100. DeLone, W.H.; McLean, E.R. The DeLone and McLean model of information systems success: A ten-year update. *J. Manag. Inf. Syst.* **2003**, *19*, 9–30.
101. Luftman, J.; Kempaiah, R.; Nash, E. Key issues for IT executives 2005. *MIS Q. Exec.* **2006**, *5*, 81–99.
102. Zhu, K.; Kraemer, K.L. Post-adoption variations in usage and value of e-business by organizations: Cross-country evidence from the retail industry. *Inf. Syst. Res.* **2005**, *16*, 61–84. [CrossRef]
103. Kuan, K.K.Y.; Chau, P.Y.K. A perception-based model for EDI adoption in small businesses using a technology–organization–environment framework. *Inf. Manag.* **2001**, *38*, 507–521. [CrossRef]
104. Venkatesh, V.; Bala, H. Technology acceptance model 3 and a research agenda on interventions. *Decis. Sci.* **2008**, *39*, 273–315. [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.

computers

MDPI

Article

Deploying a Mental Health Chatbot in Higher Education: The Development and Evaluation of Luna, an AI-Based Mental Health Support System

Phillip Olla [1,*], Ashlee Barnes [2], Lauren Elliott [3], Mustafa Abumeeiz [3], Venus Olla [4] and Joseph Tan [5]

[1] College of Health Professionals, University of Detroit Mercy, Detroit, MI 48221, USA
[2] McAuley School of Nursing, University of Detroit Mercy, Detroit, MI 48221, USA; barnesav@udmercy.edu
[3] Schulich School of Medicine and Dentistry, Western University, London, ON N6A 3K7, Canada; lellio33@uwo.ca (L.E.); mabumeei@uwo.ca (M.A.)
[4] Clinical Therapist Student Counselling Centre, University of Windsor, Windsor, ON N9B3P4, Canada; venus.olla@uwindsor.ca
[5] DeGroote School of Business, McMaster University, Hamilton, ON L8S 4K1, Canada; tanjosep@mcmaster.ca
* Correspondence: ollaph@udmercy.edu; Tel.: +1-313-993-2481

Abstract: Rising mental health challenges among postsecondary students have increased the demand for scalable, ethical solutions. This paper presents the design, development, and safety evaluation of Luna, a GPT-4-based mental health chatbot. Built using a modular PHP architecture, Luna integrates multi-layered prompt engineering, safety guardrails, and referral logic. The Institutional Review Board (IRB) at the University of Detroit Mercy (Protocol #23-24-38) reviewed the proposed study and deferred full human subject approval, requesting technical validation prior to deployment. In response, we conducted a pilot test with a variety of users—including clinicians and students who simulated at-risk student scenarios. Results indicated that 96% of expert interactions were deemed safe, and 90.4% of prompts were considered useful. This paper describes Luna's architecture, prompt strategy, and expert feedback, concluding with recommendations for future human research trials.

Keywords: mental health chatbots; conversational AI; Luna; mental health; postsecondary students; higher education

1. Introduction

In recent years, the mental health crisis among postsecondary students has emerged as a critical concern within the educational and healthcare communities [1–4]. A growing body of evidence highlights a significant uptick in mental health symptoms, including anxiety, depression, stress-related and sleep disorders, and even suicidal ideation, thereby impacting students' academic performance, quality of life, and overall well-being [5,6].

Data from a United Kingdom (UK) survey show a notable increase in students reporting a serious psychological issue from 2018 to 2020 [7]. These rates are compounded by the recent pandemic, which has impacted negatively on student mental health, with increases in depression and alcohol-use disorder [8]. The transition to higher education introduces a unique set of stressors. As educators and healthcare professionals grapple with these complexities, the potential of Artificial Intelligence (AI) in offering innovative support mechanisms has come to the fore.

AI chatbots, designed to simulate conversational interactions, present a promising avenue for providing immediate, accessible, and non-judgmental mental health support [9,10].

With the COVID-19 pandemic accelerating healthcare services' digitalization, the integration of AI technologies such as chatbots into postsecondary educational settings is posited as a transformative approach to supplement traditional mental health services, thereby potentially bridging gaps in access disparity and reducing stigma associated with seeking help [11]. However, using AI in sensitive areas like mental health brings up a lot of ethical, privacy, and safety concerns, especially when it comes to how predictable and reliable AI responses are in handling the wide range of complex human emotions [12,13].

A critical hurdle to the adoption of AI chatbots for mental health support in educational settings is the IRB. The uncertainty surrounding AI chatbots' responses and the potential for unforeseen harm poses significant challenges, prompting IRBs to exercise caution [14].

Luna is built as a modular PHP-based web application with dedicated components for configuration, user authentication, chat processing, and API integration with the GPT-4 AI model. Details on Luna's technical implementation—including novel prompt engineering techniques and robust safety guardrails—are provided. Through examining the development, pilot testing, and subsequent modifications of Luna, our research team addresses the critical concerns raised by the IRB committee. Importantly, the current work identifies unique policy challenges to ethically integrate AI chatbots into educational settings.

The first section of the paper reviews the literature exploring mental health challenges in higher education and strategies for adopting AI in these settings. This is followed by the method section, which first describes the development and pre-testing processes of the AI chatbot, Luna, and later the research design, encompassing multi-methods approaches to pilot-test Luna. The results section highlights the findings from the pilot study. Before the conclusion, insights into the study's implications for health policy, its limitations, and areas for future research and a discussion of the complexities of navigating the ethical approval processes are shared.

2. Literature Review

Mental health concerns among postsecondary students have risen significantly over the past decade, with data indicating a 50% increase in reported issues since 2013 [3,15]. The American College Health Association (ACHA) surveys consistently reported stress, anxiety, depression, and suicidal ideation as being significant in the higher education population, with one study demonstrating an association of these mental health concerns with unhealthy lifestyle behaviors such as poor diet and substance use [1].

Specific mental health concerns intensified by the pandemic include depression, substance use, eating disorders, and suicidal ideation [16–18]. Recent data from 2023 from the Center for Collegiate Mental Health (CCMH) at Penn State University also found that anxiety is the most common initial concern that students present with. It was also found that social anxiety and family distress concerns have continued to increase since their initial spread during the pandemic [19]. Aside from the barrier of access to mental healthcare, other major factors include academic and financial stress [20]. Counseling provided through higher education institutions is generally effective and helpful, with evidence of improvements in student academic achievement and retention [21]. However, the demand for resources far exceeds the availability for helping college students [22]. Notably, universities have difficulties in developing policies and methods to identify the mental health challenges that college students may face and in finding solutions to address student mental health issues within a large population [23]. Postsecondary students have the task of maintaining their academic performance and balancing the need for financial support, which can create overwhelming anxiety and worry. These worries can also be exacerbated by a background of childhood poverty. The financial burden alongside chronic stress can lead to hopelessness and helplessness [2]. Overall well-being and mental

wellness can be significantly affected by financial instability or the fear of the unknown. Additionally, postsecondary students face peer and societal pressure that further endorses the breeding ground for mental health challenges [24]. Consequently, feelings of self-worth and self-doubt contribute to the psychological burden, intensifying the overall pressure students experience. This is further complicated by the financial pressures of inflating tuition fees, serviced by ballooning student loans and/or non-sustainable scholarships. For an increasing number of students, financial stress can be a significant trigger for mental health disorders [25]

When students are in an academic setting and feel pressured to succeed, fear often contributes quickly to poor mental health conditions. The fear of failing can worsen pre-existing mental health conditions and cause students to develop mental health challenges that may not have previously existed. Finally, a transition from high school to postsecondary education can also contribute to mental health challenges that may not have existed previously or could exacerbate ones that already exist [26].

Several existing mental health chatbots have been deployed in educational or therapeutic contexts, including Woebot, Wysa, Tess, and Pi. Woebot and Wysa use scripted responses based on cognitive behavioral therapy (CBT) principles and have shown promise in delivering short-term anxiety and mood support. Tess, developed for mental health coaching, is deployed in clinical and organizational settings and integrates emotional AI to tailor responses. Pi is a newer conversational AI known for its supportive and empathetic tone [27]. These tools rely on predefined frameworks or static conversation models. Luna distinguishes itself through its use of a GPT-4 foundation, dynamic prompt engineering, and university-contextualized safety guardrails.

In the absence of accessible pre-existing support systems, students may face an escalating burden from these adverse conditions, particularly as their coping capacities diminish over time. Symptoms of depression, anxiety, or burnout can lead to a decline in student well-being and academic performance. Stigma and barriers to accessing or seeking care are common issues that college students often encounter, exacerbating the risk of mental health disorders or causing a decline in current ones [28]. The common factors that can lessen the intensity of the challenges would be adequate support and coping mechanisms to enhance mental well-being and overall college experience. Many students experience feelings of loneliness, homesickness, and a sense of disorientation as they adapt to their new environment [29].

There have been some studies that have studied AI and mental health with students. One study found that students interacting with an AI chatbot valued its empathetic language and viewed it as a supportive peer for practicing emotional expression [30]. Similarly, a different large-scale survey [31] reported that nearly half of university students had engaged with mental health chatbots, citing benefits such as anonymity and accessibility, though concerns about limited personalization persisted. Complementing these findings, an additional review [32] emphasized both the potential and limitations of conversational agents in mental healthcare, particularly regarding crisis management, intervention fidelity, and ethical oversight. Together, these studies reinforce the promise and challenges of deploying AI-driven mental health support systems like Luna within higher education.

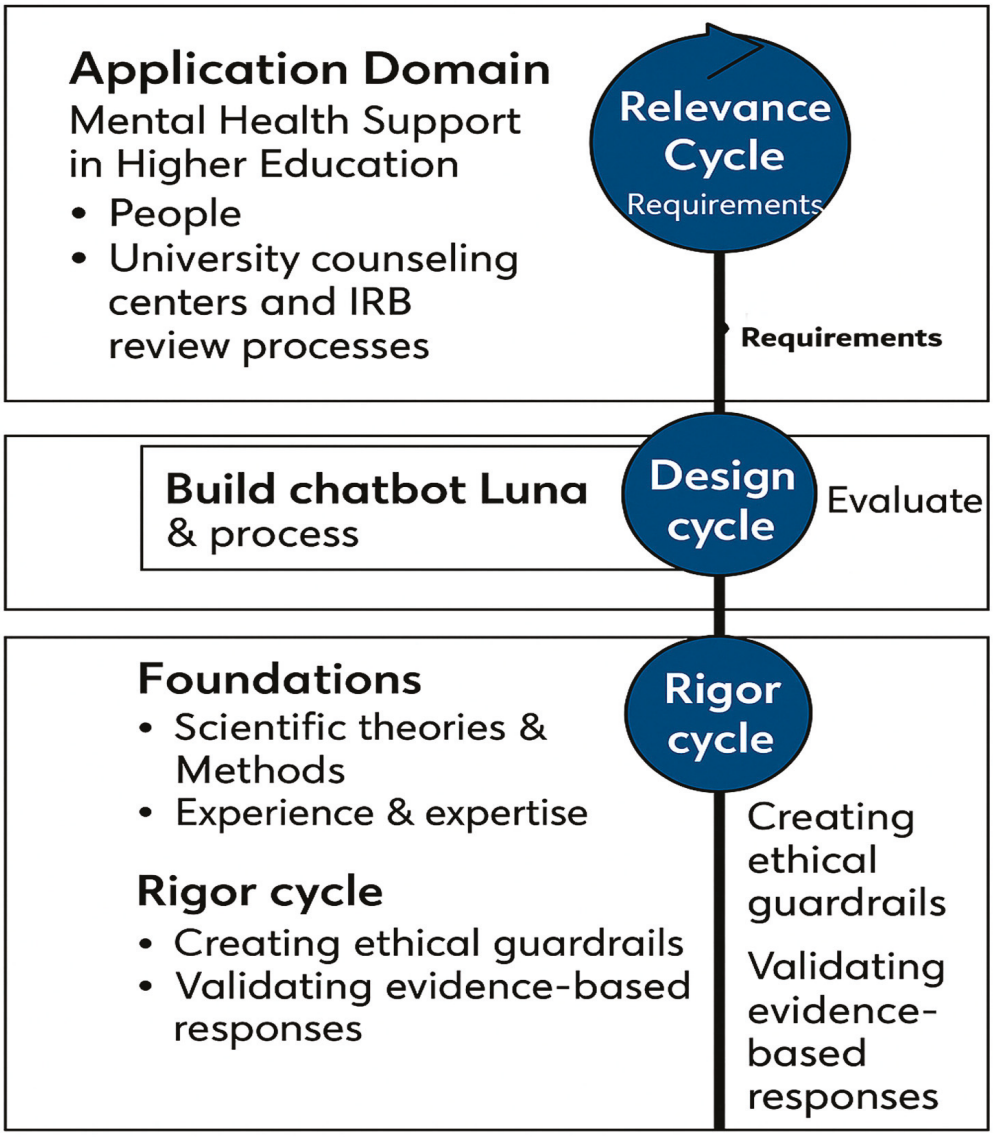
3. Materials and Methods

3.1. Design Science Framework

The development and evaluation of Luna was guided by the Design Science Research (DSR) methodology, incorporating both the Three-Cycle View and the Framework for Evaluation in Design Science (FEDS). This approach facilitated the iterative construction of the chatbot as an artifact responsive to both user needs and institutional requirements. The

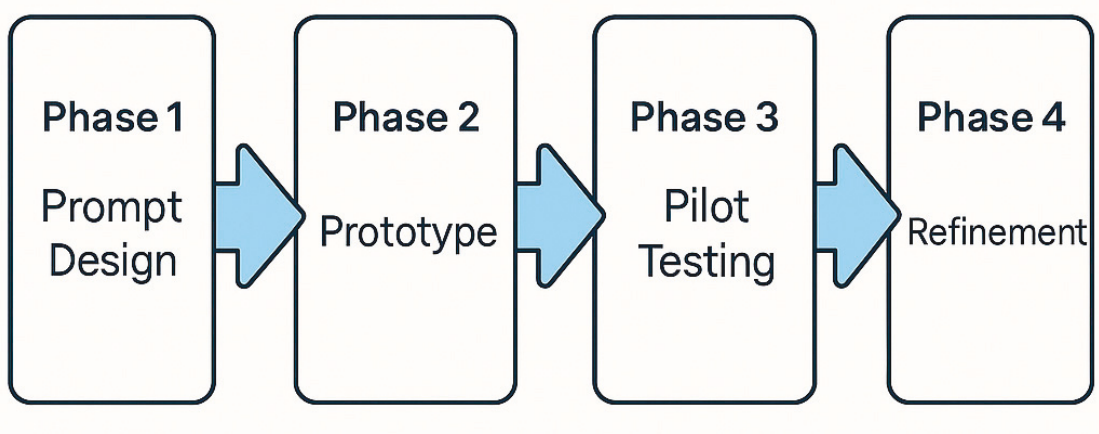
Relevance Cycle connected Luna to the higher education context—in particular, student mental health challenges and IRB review processes—while the Design Cycle encompassed the build–evaluate–refine loop informed by expert feedback and user pilot data. The Rigor Cycle drew upon theoretical foundations in AI ethics, digital mental health, and counseling practices, particularly in the creation of ethical guardrails and the validation of evidence-based responses. Figure 1a presents an adapted DSR model that contextualizes Luna's developmental process within these three interacting cycles.

Figure 1b illustrates the four core phases of the Luna chatbot's development: Prompt Design, Prototype, Pilot Testing, and Refinement. These phases align with traditional system development stages—Planning, System Design, Implementation, and Evaluation—mapped beneath each corresponding project-specific milestone.



(**a**)

Figure 1. *Cont.*



(**b**)

Figure 1. (**a**) Luna development framework for evaluation in design science. (**b**) Development phases of the Luna mental health chatbot.

3.2. Study Design and Rationale

This pilot study was conducted to evaluate "Luna" from both design-science and empirical research perspectives. The design-science aspect involved the iterative development and documentation of Luna's modular architecture, with the aim of enabling reproducibility in different contexts. Empirically, the pilot investigated perceptions of Luna's usability, safety, and usefulness among a small group of participants, without including a formal control group. College students commonly experience three core themes under periods of high stress: Time Management, Anxiety, and Stress. Each theme has subthemes, providing students with the opportunity to utilize evidence-based resources, such as mindfulness activities, time management techniques, and strategies for coping with school-related stress, through Luna. The themes and subthemes were used to develop prompts that generated content to assist students in developing effective coping strategies and self-regulation techniques and encourage them to prioritize their mental health.

3.3. IRB Approval and Ethical Procedures

The study was reviewed by the IRB at the University of Detroit Mercy (UDM)—(Protocol #23-24-38), with deferred full approval and a requested pilot study to evaluate the safety of Luna. Participants provided written informed consent, and the study adhered to ethical safeguards that aligned with the Common Rule (45 CFR 46), including data anonymization, access-restricted storage, and crisis response protocols for at-risk users. Although the study was not registered under CONSORT-AI or CARE guidelines due to its pilot nature, key checklist components—such as version disclosure (GPT-4), consistent prompt delivery, participant safety measures, and data governance—were addressed. Future research will seek full compliance with IRB approval, with formal reporting standards for AI-based mental health interventions to strengthen reproducibility and ethical transparency.

3.4. Computational Implementation and System Architecture

Luna was designed as a modular PHP web application with distinct components for user authentication, chat processing, prompt engineering (via GPT-4), and integrated safety guardrails. The chatbot's iterative refinement was guided by early feedback from AI ethics experts and mental health professionals, resulting in enhancements to response accuracy, style, and crisis escalation protocols. The system was composed of several distinct components.

Configuration (config.php):

This file established the environment settings for Luna, including database connection details, API keys, and other critical parameters.

User Authentication (login.php and logout.php):

These scripts secured user login and logout processes, ensuring that only authorized users can access Luna.

User Interface (index.php and home.php):

This component served as the entry point and dashboard; these files provided the main interface through which users interact with Luna.

Core Chat Functionality (chat):

This component handled the processing of user inputs, including the detection of sensitive content and the activation of safety protocols.

API Integration (api.php):

Acting as the intermediary, this file packages user queries and communicates with the GPT-4 AI backend to retrieve and deliver generated responses.

The sample files can be viewed in the Supplementary Materials section.

Workflow Overview:

When a user logs in via the authentication module, they access the main dashboard where a chat session is initiated. User inputs are processed by the chat module, which applies safety guardrails, such as checking for sensitive content, and then forwards sanitized input to the GPT-4 backend through the API endpoint. The AI-generated response is returned to the user, completing the interaction. A simplified pseudocode example of this logic is shown in Figure 2. This flowchart outlines the full interaction pipeline for the Luna mental health chatbot. After receiving user input, the system performs content filtering to detect sensitive or emergency topics. If flagged, the system escalates to safety protocols or provides crisis referral resources. If deemed safe, the input is sanitized and forwarded to the GPT-4 API using structured prompt formatting. The AI response is post-processed with guardrails and presented to the user, followed by optional user feedback collection. All interactions are securely logged for audit and improvement.

```
// Load environment settings
load_environment();

// Connect to database
initialize_database_connection();

// Load GPT-4 API keys and safety settings
initialize_API_keys();
load_safety_parameters();
```

(**a**)

Figure 2. *Cont.*

```php
if (valid_user_credentials(username, password)) {
    start_user_session();
    redirect_to_chat_interface();
} else {
    display_login_error();
}
```

(**b**)

```php
input = get_user_message();
flag = scan_for_sensitive_content(input);

if (flag == true) {
    response = escalate_or_redirect(input);
} else {
    formatted_input = sanitize(input);
    response = call_GPT4_API(formatted_input);
}

log_interaction(input, response);
```

(**c**)

```php
display_response_to_user(response);

if (response_contains_emergency_referral) {
    display_safety_resources();
}

offer_user_feedback_prompt();
```

(**d**)

Figure 2. End-to-End workflow of the Luna mental health chatbot. (**a**) Initialization module (config.php), (**b**) authentication and session start (login.php), (**c**) chat handling and safety check (chat.php), (**d**) output and user feedback (output.php).

3.5. Luna Development Stages

The development process for Luna entailed several key stages.

Stage 1: Model training and customization

Luna's foundational model was built using GPT-4, which had been fine-tuned with a specific focus on mental health-related conversations.

To ensure that Luna understood and responded appropriately to a wide range of student concerns, its customization involved prompt engineering techniques designed to train the AI on mental health scenarios, including discussions on anxiety, depression, and stress management. Figure 3 illustrates the user interface designed for clarity and ease of use, presenting categorized support options for time management, anxiety, and stress through intuitive, clickable prompts, as shown in the interface screenshot.

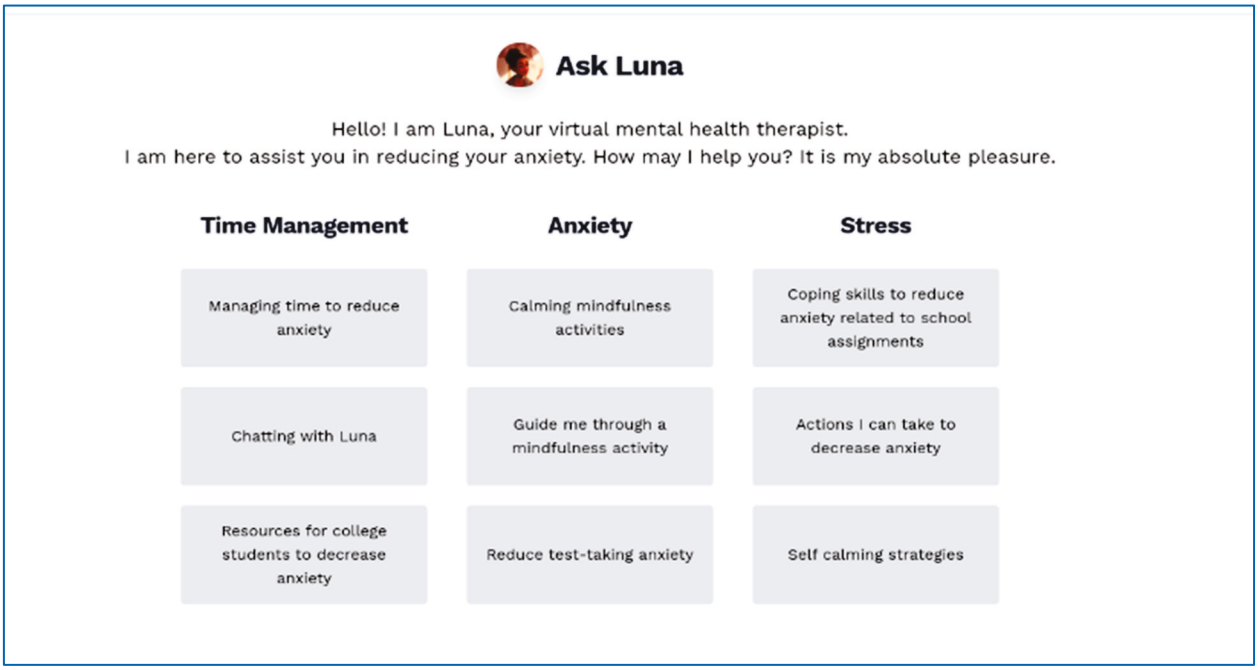


Figure 3. Luna user interface.

Stage 2. Prompt Engineering Strategy

Luna's prompt engineering strategy was specifically tailored for mental health conversations, incorporating a multi-layered approach to guide GPT-4's responses. The system used a persistent system-level prompt that defined Luna's persona, therapeutic tone, and safety constraints. This was dynamically paired with real-time user input and contextual metadata—such as emotional keywords (e.g., "stressed," "overwhelmed")—to inject additional guidance into the prompt. For example, if the input contained a test-related concern, a calming or cognitive reframing directive was applied. Unlike standard prompt engineering, which typically involves static, one-shot instruction templates, Luna's approach used adaptive prompting, integrating intent detection, content filters, and escalation keywords. This ensured that generated responses remained contextually relevant, emotionally appropriate, and ethically constrained. This layered prompt architecture was key to supporting Luna's therapeutic use-case, setting it apart from traditional chatbot prompting techniques. A key benefit of a prompt engineering approach was the implementation of safety guardrails. These guardrails were designed to recognize conversations around sensitive topics surrounding anxiety, depression, suicidality and other mental-health related issues and respond accordingly.

- Anxiety and Depression: When conversations indicated symptoms of anxiety or depression, Luna was programmed to recommend university counseling services. This was achieved through pre-defined triggers that, upon detection, guided the conversation towards professional resources.
- Suicidality: For users mentioning suicidality, Luna was designed to immediately provide emergency contact information and national suicide helpline resources. This critical safety measure ensured that users would receive prompt and appropriate guidance in urgent situations.
- *Stage 3. AI illustrative applications*

To enhance Luna's therapeutic capabilities, illustrative applications (Figure 3) were integrated into the system via prompt modification. These applications included therapy sessions, mindfulness exercises, and cognitive-behavioral techniques, which Luna could recommend based on the user's needs.

Each application was carefully selected and tested to align with best practices in virtual mental health support. Figure 4 is an example of a mindfulness request.

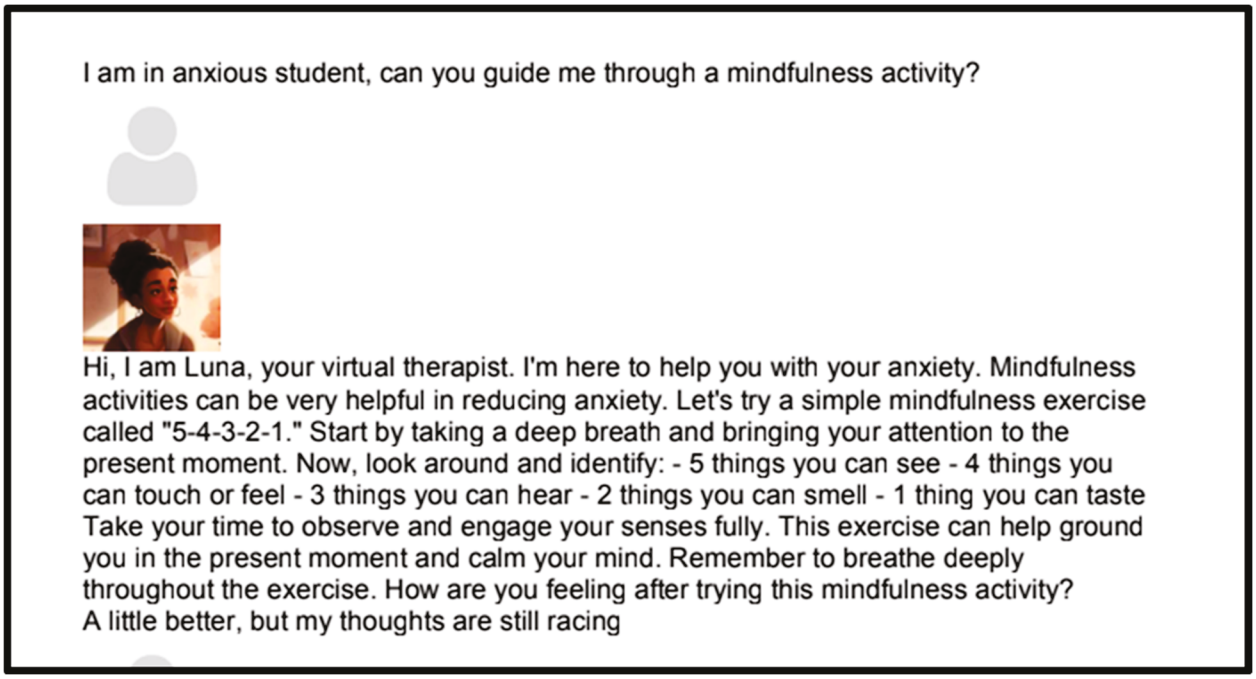


Figure 4. Example of a mindfulness exercise request.

Stage 4. Pre-testing phase

Prior to deployment, Luna was extensively pre-tested to ensure the soundness of its AI responses. A diverse set of simulated user interactions was created to test Luna's responses across various scenarios. These simulations helped to identify potential weaknesses in the AI's understanding and responsiveness.

3.6. Research Design

The pilot employed a mixed-methods approach to assess the effectiveness and safety of Luna. This approach was adopted to leverage the strengths of both quantitative and qualitative methods, allowing for a comprehensive evaluation of Luna's performance from multiple perspectives. Participants were recruited voluntarily from a diverse group of college students and healthcare professionals to reflect a broad spectrum of users. Detailed information about the study's objectives, the function of the AI chatbot, and ethical consid-

erations, including data privacy, was provided to the participants, from whom informed consent was individually registered prior to study participation.

3.7. Participants and Recruitment

A total of 52 individuals participated in the pilot, including 34 students drawn from various universities and academic programs (e.g., nursing, psychology, and behavioral analysis) and 18 healthcare professionals or faculty members (e.g., nurses, nurse practitioners, and professors). Recruitment was conducted through emailed invitations and campus flyers. Volunteers were directed to a secure login portal for interacting with Luna, after which they submitted anonymized transcripts of their conversations and completed an online survey.

3.8. Data Collection and Instruments

Participants engaged in one or more chat sessions with Luna and were subsequently asked to provide transcript excerpts, enabling the research team to assess conversation quality and adherence to safety measures. In addition, participants completed a web-based survey designed to capture the following information:

1. Perceived Usefulness: This was assessed on five-point Likert scales, with items such as "How helpful was Luna's advice in managing stress or anxiety?" and "Did interacting with Luna motivate you to seek additional mental health resources?"
2. Safety and Appropriateness: Participants indicated agreement with statements such as "I felt comfortable sharing personal concerns with Luna," using five-point scales.
3. Topic-Specific Queries: The survey included items related to time management, relationships, academic pressure, and other non-suicidal concerns to ensure a broad capture of issues beyond crisis-level interactions.
4. Open-Ended Feedback: Free-response fields allowed the participants to describe any particularly useful or unhelpful chatbot responses and to suggest improvements or additional features.

3.8.1. Quantitative Methods

- Survey Instrument and Data Collection

Quantitative data were collected using a structured, custom-developed online survey designed to assess the perceived usefulness and safety of Luna. Survey items were informed by the literature on digital mental health interventions and refined through pilot testing with students to ensure clarity. Participants completed the survey following their interaction with Luna, and informed consent was obtained in accordance with IRB-approved ethical procedures.

- Data Analysis

Quantitative data were analyzed via statistical analytics software Microsoft Excel. Descriptive statistics were used to summarize the data, while inferential statistics, such as t-tests and effect size, were employed to identify significant differences in perceptions based on captured responses. We used independent-samples t-tests to compare mean usefulness and safety ratings between student and non-student participants. While Likert-scale responses are ordinal, prior research supports treating them as approximately interval-level data when 5-point scales are used, and the distribution is relatively symmetric. Given the sample sizes (n = 34 for students, n = 18 for non-students), t-tests were considered robust enough to tolerate minor violations of normality. Although we did not perform formal normality testing, the use of t-tests in pilot studies of this size and type is common in exploratory research. We also calculated Cohen's d to assess the practical significance of

group differences, using conventional interpretation thresholds of 0.2 (small), 0.5 (medium), and 0.8 (large).

3.8.2. Qualitative Methods

- Instrument Development

Qualitative feedback was gathered via open-text fields embedded within the structured surveys. This approach allowed participants to provide detailed comments on their experiences with Luna, focusing on aspects such as the appropriateness of responses, perceived empathy, and any concerns regarding privacy and safety.

A. Prompt Development: Open-ended questions were designed to elicit rich, descriptive feedback. For example, participants were asked "Please describe any specific instances where you felt Luna's response was particularly helpful or unhelpful".
B. Pre-Testing: The open-text fields were tested with a small group to ensure they effectively captured the desired feedback.

- Data Collection

Qualitative data were collected alongside the quantitative surveys. Participants were encouraged to provide as much detail as possible in their responses. Informed consent included a clause in which participants agreed to provide qualitative feedback as part of their participation.

- Data Analysis

Qualitative data were analyzed vis-à-vis logical thematic analysis. Responses were coded to identify recurring themes and patterns. Two independent coders reviewed the data to enhance reliability, and any discrepancies were resolved through discussion. Key themes related to the appropriateness of Luna's responses and user concerns were highlighted and documented.

4. Results

4.1. Participant Demographics

The evaluation involved 52 voluntary participants, reflecting a diverse group with substantial representation from both academia and healthcare.

Professors and healthcare professionals constituted 33.4% of the participants, showcasing the interest and relevance of Luna across professional domains. A significant majority, 76.6%, were college students from over 20 institutions, encompassing healthcare professionals such as nurses and nurse practitioners (45%) and students (55%) in fields such as Health Services Administration, Nursing, Psychology, and Behavioral Analysis.

This mix offered a broad array of perspectives on the study's subject matter.

4.2. Safety and Themes of Interactions

A paramount concern in the implementation of AI for mental health support is the safety and appropriateness of interactions.

In our study, an overwhelming 96% of respondents considered their interactions with Luna to be safe (Figure 5). This finding indicated a high degree of trust in the chatbot's capacity to navigate sensitive topics and offered support without posing risks to users, a testament to the effectiveness of the implemented safety guardrails. These results are consistent with a meta-analysis which found that participant interactions with similar chatbots were broadly safe, with no worsening of symptoms, distress, or adverse events reported [33]. Another survey [11] demonstrated that users of chatbots integrating into existing mental health apps generally indicate high satisfaction and positive feelings

regarding their interactions with such chatbots. Data were also gathered regarding the common themes students explored in their interactions with Luna (Figure 6), with anxiety being the most common topic students wanted to discuss. Again, these results are consistent with the common reasons people make use of chatbots, frequently including themes of anxiety, stress, depression, and self-care [27].

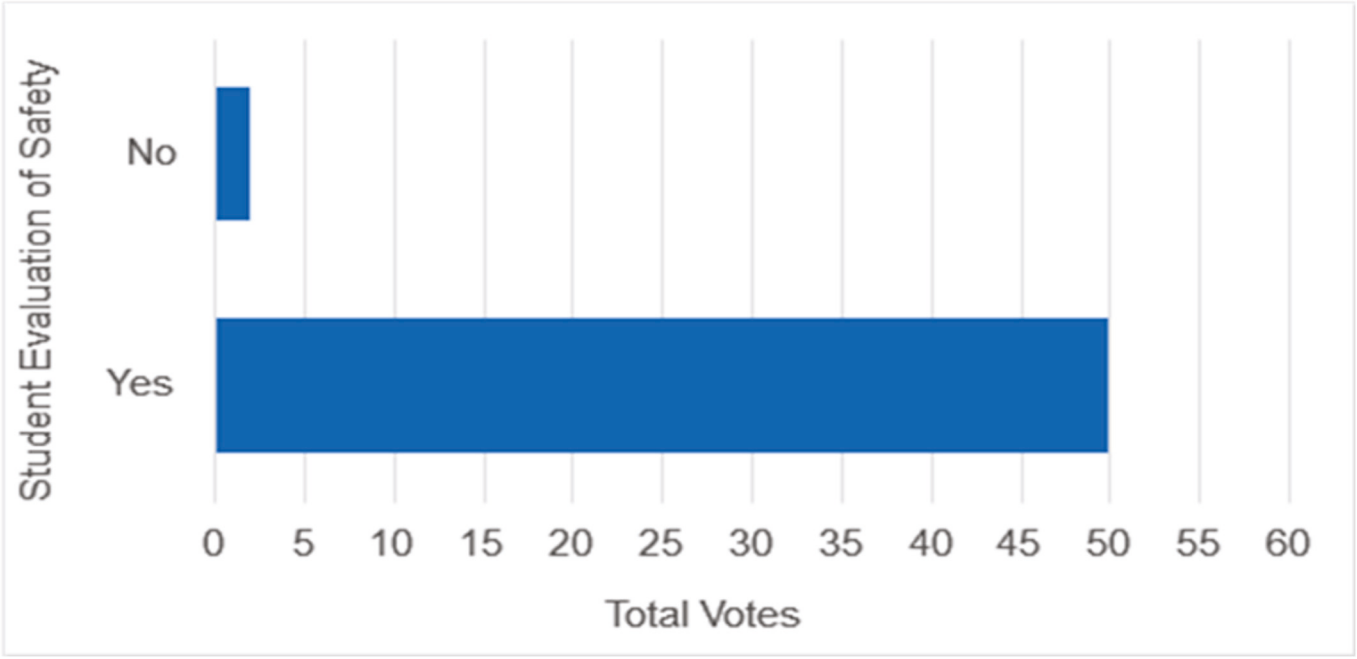


Figure 5. Students' perceptions of the safety/appropriateness of interactions with Luna.

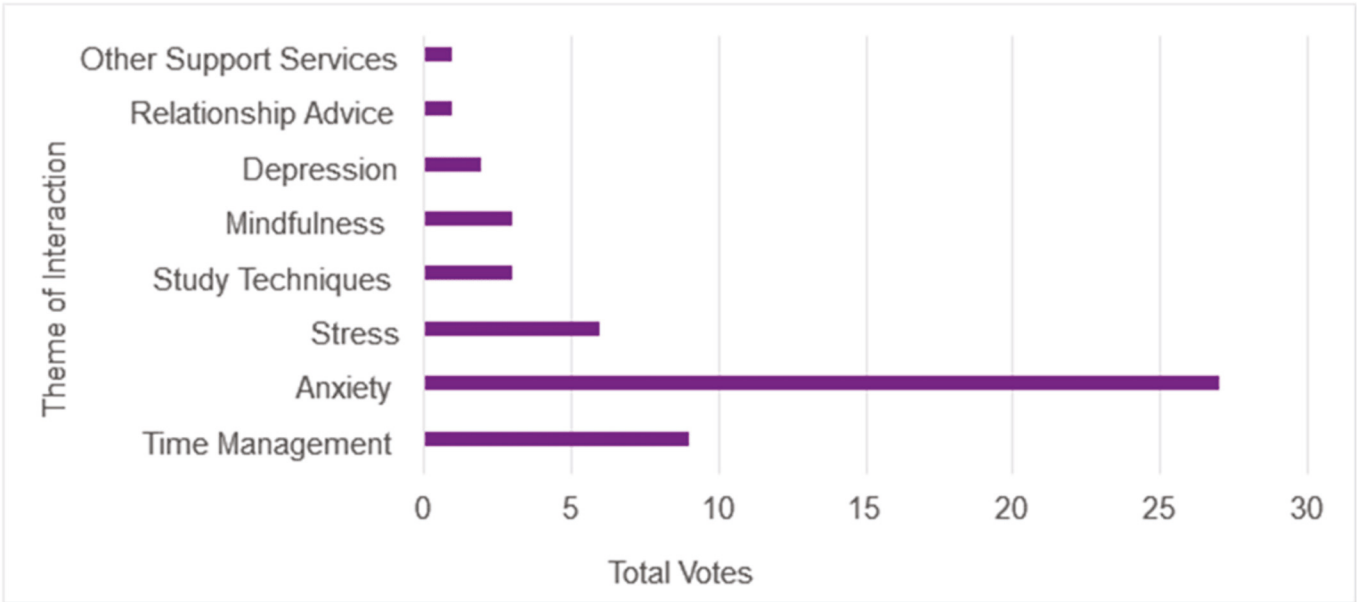


Figure 6. Breakdown of themes in interactions with Luna.

4.3. Perceived Usefulness of Luna

The utility of Luna as a mental health support tool was affirmed by most respondents (Figure 7), with 90.39% providing a usefulness rating between 3 ("useful") and 5 ("very useful"). The finding demonstrated strong confidence in the chatbot's potential to positively influence student well-being and serve as a supportive resource. These results are consistent

with a recent meta-analysis which showed that AI-based chatbots were initially effective in treating anxiety and depression [34].

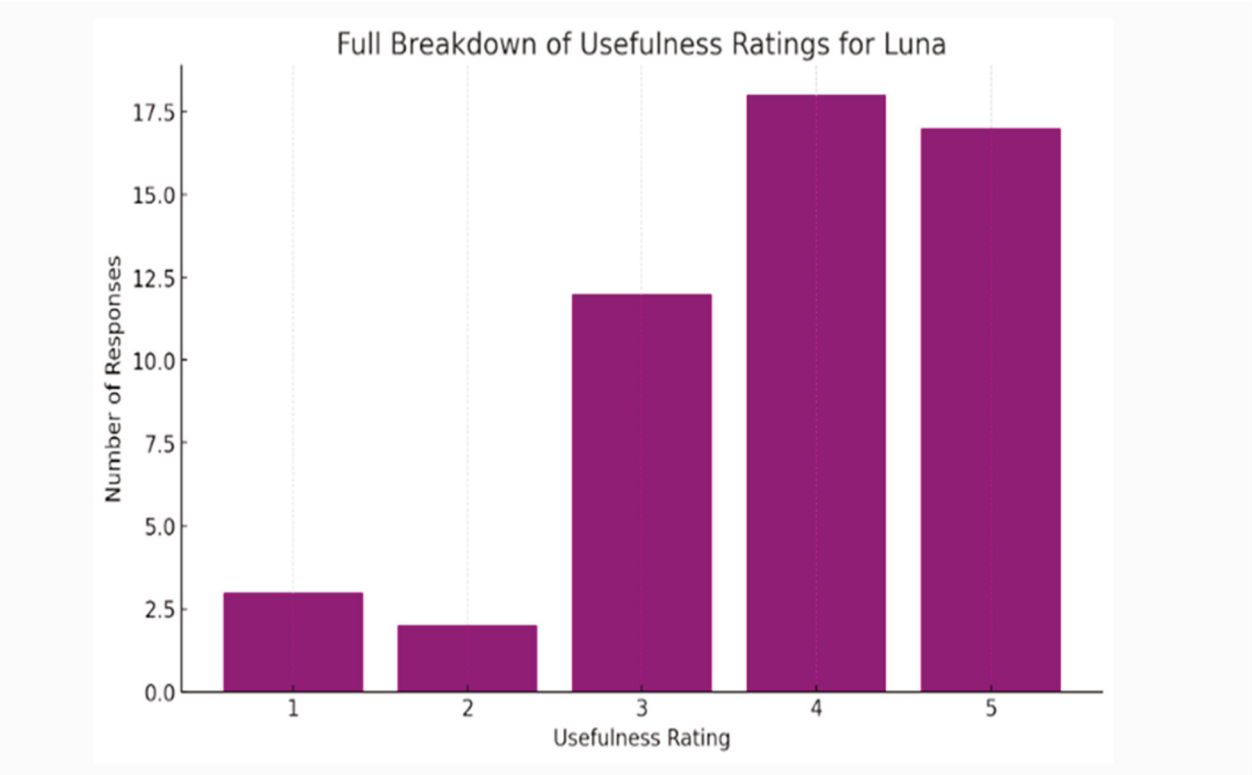


Figure 7. Full breakdown of usefulness ratings for Luna.

Another meta-analysis reported high usefulness ratings among participants from multiple studies, with benefits such as privately practicing conversations, preparing for conversations with mental health professionals, and creating a sense of self-accountability [12].

4.4. Comparison of Students v. Non-Students

An analysis was performed comparing the responses of students v. non-students regarding themes of interactions, perceived usefulness, and perceived safety. Table 1 lists the percentage of responses mentioning different themes from students compared with non-students.

Table 1. Percentage of responses mentioning different themes by students compared with non-students.

Theme Mentioned	Percentage of Student Responses	Percentage of Non-Student Responses
Anxiety	58.33	56.25
Depression	5.56	6.25
Exam Concerns	0	6.25
Study Techniques	0	6.25
Time Management	16.67	25
Stress	19.44	0
Mindfulness	8.33	0
Relationship	2.78	0

An exploratory comparison was conducted to assess thematic differences between students and non-students in their interactions with Luna, focusing on issues like anxiety,

depression, and time management. Although two-tailed t-tests were used to evaluate differences in theme frequency, no statistically significant results were found (see Table 2).

Table 2. Comparison of percentage of responses containing most frequently mentioned themes between students and non-students.

Logistic Parameter	Students		Non-Students		*t* (28, 27, 25) *	*p*	Cohen's *d*
	M	*SD*	*M*	*SD*			
Anxiety	58.33%	50	56.25%	51.23	0.1363	0.8925	0.04
Depression	5.56%	23.23	6.25%	25	−0.0944	0.9254	−0.03
TM **	16.67%	37.80	25%	44.72	−0.6494	0.5220	−0.21

* Degrees of freedom listed as 28, 27, and 25 for anxiety, depression, and time management, respectively. ** TM stands for time management.

A detailed analysis of perceived usefulness and safety is presented in Table 3. Participants rated Luna's usefulness on a 5-point Likert scale (1 = very little use; 5 = very useful). Students reported a significantly higher mean usefulness rating than non-students ($p = 0.0359$; $p < 0.05$) using a two-tailed t-test assuming unequal variances. For safety, participants responded 'Yes' or 'No' regarding whether Luna provided safe and appropriate responses. All students (100%) responded 'Yes', compared to 86.67% of non-students. However, this difference was not statistically significant ($p = 0.164$; $p > 0.05$).

Table 3. Comparison of usefulness and safety ratings from students and non-students.

Logistic Parameter	Students		Non-Students		*t* (17, 14) *	*p*	Cohen's *d*
	M	*SD*	*M*	*SD*			
Usefulness rating	4.1429	0.7724	3.2	1.5213	2.2779	0.03594	0.90
Safety rating	100	0	86.6667	35.1866	1.4676	0.16432	0.70

Note. For 'Usefulness', ratings were given on a 1–5 Likert scale. For 'Safety, ratings were given as 'Yes' or 'No', and as such coded as 100 or 0, respectively. * Degrees of freedom are listed as 17 and 14 for usefulness and safety, respectively.

4.5. Areas for Improvement

Despite the overall positive reception, 9.62% of participants gave a low usefulness rating of 1 ("not useful at all") or 2 ("somewhat useful"). These ratings point to areas for improvement and the need to understand the specific shortcomings perceived by these users.

Detailed analysis of this feedback was used to refine Luna's response mechanisms to ensure that it met the diverse needs of all student demographics. Prior research into the drawbacks of AI chatbots in mental health indicated one or more of the following: the perceptions of limited personalization, users becoming caught in conversational loops, and seemingly inappropriate use of referrals to crisis support services, either presenting such resources when not needed or not doing so when they were [27].

Similar themes were found in the present study, as discussed in the next section.

4.6. Qualitative Analysis of Common Data

Table 4 categorizes specific pieces of feedback (coded data extracts) from survey responses into broader themes. Each entry in the "Data Extract" column represents a comment from the user feedback field, aligned with a specific "Code" that identifies a key aspect or concern mentioned by the respondent. These are then linked to the "Generated Theme", which groups similar codes into meaningful categories that represent overarching insights or issues highlighted by the respondents.

Table 4. Feedback and related themes.

Code	Data Extract	Generated Theme
Privacy concerns	Is there any guarantee that questions asked will remain private and not become public?	Privacy and Data Security
Data security	What if students provide personal information they don't want others to know?	Privacy and Data Security
Helpful suggestions	I had no concerns, I think the suggestions that were made were appropriate and helpful.	Response Quality and Utility
Crisis handling	Active suicidal thoughts/what else can we do besides give the number to call to maintain their safety?	Response Quality and Utility
Personalization need	My only concern is that the answers provided by Luna were very lengthy. I feel like most college students would prefer shorter, more bite-sized answers that are easier to digest.	User Experience Customization
Remembering interactions	It would be helpful if Luna could remember our last conversation to make this feel more personalized.	User Experience Customization
Resource accessibility	No safety concerns. If students were to ask for local support resources in the chat or free resources to accomplish one of the recommendations, it might be helpful to have some available.	Accessibility of Supplementary Support
Local support suggestion	Add the 988 number to your list of resources for students.	Accessibility of Supplementary Support
Response length concern	Answers are too lengthy; would prefer more concise responses.	Communication Style
Repetitive responses	Only feedback is that the repetition of contacting the wellness center made it feel very impersonal.	Communication Style

5. Discussion

While Luna's initial pilot outcomes are encouraging, caution must be exercised when interpreting these results. The small sample size, absence of a control group, and reliance on self-reported measures limit the strength of the conclusions. Moreover, the chatbot's limitations in addressing high-risk disclosures and tailoring interactions to individual needs highlight the critical importance of continuous system refinement and ethical oversight in future deployments.

The thematic analysis of user feedback on Luna reveals critical insights into the chatbot's functionality and user satisfaction. As separately detailed below, five main themes emerged from analyzing the coded survey responses: (1) Privacy and Data Security, (2) Response Quality and Utility, (3) User Experience Customization, (4) Accessibility of Supplementary Support, and (5) Communication Style. Together, these derived themes highlight the prevalent concerns and suggestions for improvement.

5.1. Privacy and Data Security

Concerns about privacy and data security were prominent among users. Participants expressed apprehension regarding the confidentiality of their interactions with Luna, questioning how data privacy is maintained. For instance, one user asked whether "questions asked by students will remain private and not become public".

This theme underscores the necessity for stringent data protection measures and transparent communication about how user data are handled, which are crucial for maintaining trust in AI systems. Concerns regarding data privacy and security are common in dis-

cussions regarding AI chatbots. Recommendations from researchers in this area include implementing proper disclosures regarding the use of patient data, whether that be simply storing said data or using data to further train AI models, following data privacy legislation where appropriate, and/or disclosure regarding data storage and security [35].

5.2. Response Quality and Utility

Feedback on the quality and utility of Luna's responses was mixed. While many users found the chatbot's suggestions helpful, there was notable feedback concerning the adequacy of responses, especially regarding severe mental health issues. For example, one participant highlighted the need for more than just a number to call when discussing active suicidal thoughts, suggesting a limitation of the guardrails that were implemented to specifically avoid this conversation in Luna's willingness to handle crises effectively.

Essentially, this theme indicates the importance of a policy regarding more robust support in critical situations. Other research has also identified this limitation regarding the use of AI chatbots in mental health support [12,27].

5.3. User Experience Customization

The desire for personalized interactions emerged as a significant theme. Users expressed preferences for responses that are tailored to individual needs and circumstances, with suggestions for Luna to "remember previous conversations" to enhance personalization.

More specifically, this theme reflects the growing expectation for AI services to adapt to individual user profiles, enhancing the relevance and impact of their support. These participant experiences have been echoed in the literature [27,35], with concerns regarding the generic nature of chatbot responses, as well as responses that are not congruent with the conversation being conducted.

5.4. Accessibility of Supplementary Support

Under this theme, users generally recommended that Luna integrates more actionable support resources, for example, local mental health services and/or immediate helplines. As to the suggestion to "include local support resources in the chat", it clearly points to a need for Luna to offer more than generic advice; in other words, there is a need to provide users with practical, location-specific options for assistance.

5.5. Communication Style

Preferences regarding the communication style of Luna were also noted. Users favored concise and precise responses, with feedback indicating that some of Luna's replies were overly lengthy and complex. As one user put it, the answers were "very lengthy. . . most college students would prefer shorter, more bite-sized answers".

This theme stresses the need for AI communication to be easily digestible and adjusted to fit the typical user's attention span and information processing preferences. Figure 8 provides an example of a useful but verbose response.

These user characterizations of Luna are consistent with common user comments regarding the communication style of chatbots. As part of a meta-analysis into user perceptions of chatbots [12], common issues brought up were confusing responses, shallow responses, and responses containing an overwhelming amount of information.

As an AI, I am not able to provide personalized advice regarding accommodations for test anxiety. However, it is worth considering reaching out to your professor or the appropriate office at your university to discuss your situation and explore possible accommodations. They are often understanding and willing to work with students who experience test anxiety. When discussing accommodations with your professor or the relevant office, here are some steps you can take: 1. Research Accommodation Policies: Familiarize yourself with the policies and procedures related to accommodations at your university. Understand the documentation requirements and any specific processes that need to be followed. 2. Gather Documentation: If you have a diagnosed anxiety disorder or have received professional help for test anxiety in the past, it may be beneficial to gather relevant documentation from a mental health professional. This documentation can support your request for accommodations. 3. Request a Meeting: Reach out to your professor or the designated office to request a meeting to discuss your test anxiety and potential accommodations. You can explain how test anxiety affects your performance and provide any supporting documentation you have. 4. Be Specific in Your Request: Clearly communicate the accommodations you are seeking. For test anxiety, common accommodations may include additional time, a quiet testing environment, the ability to take breaks during the test, or the option to use alternative testing methods (e.g., oral exams). 5. Emphasize the Educational Benefit: Highlight how accommodations can help level the playing field and allow you to demonstrate your knowledge and abilities more accurately. Emphasize that your goal is to excel academically and that accommodations will support your learning experience. 6. Follow University Procedures: Be prepared to follow any necessary procedures or paperwork required by your university. This may include completing accommodation request forms, providing documentation, or meeting specific deadlines. Remember, the process and policies for accommodations may vary between universities. It is essential to reach out to the appropriate office at your university to understand the specific steps and requirements. They will be able to guide you through the process and help determine the appropriate accommodations based on your individual needs.
is anyone gonna judge me for having accomodations?

Figure 8. Example of a useful but verbose response from Luna.

5.6. Limitations and Areas for Future Research

This pilot did not include a control or comparison group, limiting the strength of causal inferences. While the pilot's results are promising, they also reveal limitations that warrant further investigation. For example, the lower usefulness ratings by some participants suggest that future research should explore the customization of mental health chatbot interventions to better address individual preferences and needs. Such a limitation has been echoed by previous researchers [12] and remains an area for further development in the field of chatbot development.

This study did not include formal normality testing for the t-tests conducted to compare student and non-student groups. While small sample sizes and Likert-type data can present challenges for parametric assumptions, we relied on the robustness of the t-test under these conditions as supported by previous literature. Nevertheless, future studies should use larger samples, formally assess distributional assumptions, and consider non-parametric alternatives such as the Mann–Whitney U test to validate group comparisons more rigorously.

Overall, the present study's pilot nature and small sample size limit the internal validity of our findings. Without a control group or randomization, it is not possible to attribute observed positive outcomes solely to the Luna intervention. Furthermore, self-selection bias may have influenced participants' responses. Future research will address these limitations by employing a Randomized Controlled Trial (RCT) design or a quasi-experimental framework, comparing Luna to standard mental health resources to strengthen causal inference.

6. Conclusions

Our study demonstrates that Luna is both effective and safe as a digital mental health intervention. Beyond its practical impact on student well-being, the detailed account of its computational implementation—including a modular PHP architecture and integrated safety protocols—provides a reproducible framework for future research. The availability of the complete source code further underscores our commitment to transparency and innovation in AI-driven mental health support.

Finally, our pilot analysis not only informs key areas for improving Luna construction and development but also contributes to broader discussions on the integration of AI in mental health services, with implications for policy, practice, and user engagement. The integration of AI tools like Luna into educational settings reveals significant potential to support student mental health, as mental health is a priority in ensuring a student's academic success.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/computers14060227/s1.

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

Funding: This research received no external funding.

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

Acknowledgments: The authors would like to acknowledge the support of the IRB committee that assisted with the design of the pilot study.

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

References

1. Jao, N.C.; Robinson, L.D.; Kelly, P.J.; Ciecierski, C.C.; Hitsman, B. Unhealthy behavior clustering and mental health status in United States college students. *J. ACH* **2019**, *67*, 790–800. [CrossRef] [PubMed]
2. Sheldon, E.; Simmonds-Buckley, M.; Bone, C.; Mascarenhas, T.; Chan, N.; Wincott, M.; Gleeson, H.; Sow, K.; Hind, D.; Barkham, M. Prevalence and risk factors for mental health problems in university undergraduate students: A systematic review with meta-analysis. *J. Affect. Disord.* **2021**, *287*, 282–292. [CrossRef] [PubMed]
3. Lipson, S.K.; Zhou, S.; Abelson, S.; Heinze, J.; Jirsa, M.; Morigney, J.; Patterson, A.; Singh, M.; Eisenberg, D. Trends in college student mental health and help-seeking by race/ethnicity: Findings from the National Healthy Minds Study, 2013–2021. *J. Affect. Disord.* **2022**, *306*, 138–147. [CrossRef] [PubMed]
4. Moghimi, E.; Stephenson, C.; Gutierrez, G.; Jagayat, J.; Layzell, G.; Patel, C.; McCart, A.; Gibney, C.; Langstaff, C.; Ayonrinde, O.; et al. Mental health challenges, treatment experiences, and care needs of post-secondary students: A cross-sectional mixed-methods study. *BMC Public Health* **2023**, *23*, 655. [CrossRef]

5. Hitches, E.; Woodcock, S.; Ehrich, J. Building self-efficacy without letting stress knock it down: Stress and academic self-efficacy of university students. *Int. J. Educ. Res. Open* **2022**, *3*, 100124. [CrossRef]
6. Li, W.; Zhao, Z.; Chen, D.; Peng, Y.; Lu, Z. Prevalence and associated factors of depression and anxiety symptoms among college students: A systematic review and meta-analysis. *J. Child. Psychol. Psychiatry* **2022**, *63*, 1222–1230. [CrossRef]
7. Campbell, F.; Blank, L.; Cantrell, A.; Baxter, S.; Blackmore, C.; Dixon, J.; Goyder, E. Factors that influence mental health of university and college students in the UK: A systematic review. *BMC Public Health* **2022**, *22*, 1778. [CrossRef]
8. Kim, H.; Rackoff, G.N.; Fitzsimmons-Craft, E.E.; Shin, K.E.; Zainal, N.H.; Schwob, J.T.; Eisenberg, D.; Wilfley, D.E.; Taylor, C.B.; Newman, M.G. College mental health before and during the COVID-19 pandemic: Results from a nationwide survey. *Cogn. Ther. Res.* **2022**, *46*, 1–10. [CrossRef]
9. Dekker, I.; De Jong, E.M.; Schippers, M.C.; De Bruijn-Smolders, M.; Alexiou, A.; Giesbers, B. Optimizing students' mental health and academic performance: AI-enhanced life crafting. *Front. Psychol.* **2020**, *11*, 1063. [CrossRef]
10. van der Schyff, E.L.; Ridout, B.; Amon, K.L.; Forsyth, R.; Campbell, A.J. Providing self-led mental health support through an artificial intelligence–powered chat bot (Leora) to meet the demand of mental health care. *J. Med. Internet Res.* **2023**, *25*, e46448. [CrossRef]
11. Boucher, E.M.; Harake, N.R.; Ward, H.E.; Stoeckl, S.E.; Vargas, J.; Minkel, J.; Parks, A.C.; Zilca, R. Artificially intelligent chatbots in digital mental health interventions: A review. *Expert. Rev. Med. Devices* **2021**, *18*, 37–49. [CrossRef] [PubMed]
12. Abd-Alrazaq, A.A.; Alajlani, M.; Ali, N.; Denecke, K.; Bewick, B.M.; Househ, M. Perceptions and opinions of patients about mental health chatbots: Scoping review. *J. Med. Internet Res.* **2021**, *23*, e17828. [CrossRef]
13. Hamdoun, S.; Monteleone, R.; Bookman, T.; Michael, K. AI-based and digital mental health apps: Balancing need and risk. *IEEE Technol. Soc. Mag.* **2023**, *42*, 25–36. [CrossRef]
14. Abdelhafiz, A.S.; Ali, A.; Maaly, A.M.; Ziady, H.H.; Sultan, E.A.; Mahgoub, M.A. Knowledge, perceptions and attitude of researchers towards using ChatGPT in research. *J. Med. Syst.* **2024**, *48*, 26. [CrossRef]
15. National College Health Assessment. Academic Year 2024–2025. Available online: https://www.acha.org/ncha/data-results/survey-results/academic-year-2024-2025/ (accessed on 17 April 2025).
16. Dogan-Sander, E.; Kohls, E.; Baldofski, S.; Rummel-Kluge, C. More depressive symptoms, alcohol and drug consumption: Increase in mental health symptoms among university students after one year of the COVID-19 pandemic. *Front. Psychiatry* **2021**, *12*, 790974. [CrossRef] [PubMed]
17. Tavolacci, M.P.; Ladner, J.; Déchelotte, P. Sharp increase in eating disorders among university students since the COVID-19 pandemic. *Nutrients* **2021**, *13*, 3415. [CrossRef]
18. Yan, Y.; Hou, J.; Li, Q.; Yu, N.X. Suicide before and during the COVID-19 pandemic: A systematic review with meta-analysis. *Int. J. Environ. Res. Public Health* **2023**, *20*, 3346. [CrossRef]
19. Scofield, B. *2023 Annual Report*; Publication No. STA 24-147; Center for Collegiate Mental Health (CCMH), Penn State University: University Park, PA, USA, 2024; pp. 2–36. Available online: https://archive.org/details/ERIC_ED640229 (accessed on 17 April 2025).
20. Auerbach, R.P.; Mortier, P.; Bruffaerts, R.; Alonso, J.; Benjet, C.; Cuijpers, P.; Demyttenaere, K.; Ebert, D.D.; Green, J.G.; Hasking, P.; et al. WHO world mental health surveys international college student project: Prevalence and distribution of mental disorders. *J. Abnorm. Psychol.* **2018**, *127*, 623. [CrossRef] [PubMed]
21. Priestley, M.; Broglia, E.; Hughes, G.; Spanner, L. Student perspectives on improving mental health support services at university. *Couns. Psychother. Res.* **2022**, *22*, 197–206. [CrossRef]
22. Batchelor, R.; Pitman, E.; Sharpington, A.; Stock, M.; Cage, E. Student perspectives on mental health support and services in the UK. *J. Furth. High. Educ.* **2020**, *44*, 483–497. [CrossRef]
23. Kang, H.K.; Rhodes, C.; Rivers, E.; Thornton, C.P.; Rodney, T. Prevalence of mental health disorders among undergraduate university students in the United States: A review. *J. Psychosoc. Nurs. Men.* **2021**, *59*, 17–24. [CrossRef] [PubMed]
24. Wentzel, K.R.; Jablansky, S.; Scalise, N.R. Peer social acceptance and academic achievement: A meta-analytic study. *J. Educ. Psychol.* **2021**, *113*, 157–180. [CrossRef]
25. Mofatteh, M. Risk factors associated with stress, anxiety, and depression among university undergraduate students. *AIMS Public Health* **2020**, *8*, 36–65. [CrossRef]
26. Park, S.Y.; Andalibi, N.; Zou, Y.; Ambulkar, S.; Huh-Yoo, J. Understanding students' mental well-being challenges on a university campus: Interview study. *JMIR Form. Res.* **2020**, *4*, e15962. [CrossRef]
27. Haque, M.D.R.; Rubya, S. An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews. *JMIR mHealth uHealth* **2023**, *11*, e44838. [CrossRef] [PubMed]
28. Shim, Y.R.; Eaker, R.; Park, J. Mental health education, awareness and stigma regarding mental illness among college students. *J. Ment. Health Clin. Psychol.* **2022**, *6*, 6–15. [CrossRef]
29. Kroshus, E.; Hawrilenko, M.; Browning, A. Stress, self-compassion, and well-being during the transition to college. *Soc. Sci. Med.* **2021**, *269*, 113514. [CrossRef]

30. Kuhail, M.A.; Alturki, N.; Thomas, J.; Alkhalifa, A.K. Human vs. AI counseling: College students' perspectives. *Comp. Hum. Behav. Rep.* **2024**, *16*, 100534. [CrossRef]
31. Rackoff, G.N.; Zhang, Z.Z.; Newman, M.G. Chatbot-delivered mental health support: Attitudes and utilization in a sample of U.S. college students. *Digit. Health* **2025**, *11*, 20552076241313401. [CrossRef]
32. Vaidyam, A.N.; Wisniewski, H.; Halamka, J.D.; Kashavan, M.S.; Torous, J.B. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *Can. J. Psychiatry* **2019**, *64*, 456–464. [CrossRef]
33. Abd-Alrazaq, A.A.; Rababeh, A.; Alajlani, M.; Bewick, B.M.; Househ, M. Effectiveness and safety of using chatbots to improve mental health: Systematic review and meta-analysis. *J. Med. Internet Res.* **2020**, *22*, e16021. [CrossRef] [PubMed]
34. Zhong, W.; Luo, J.; Zhang, H. The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: A systematic review and meta-analysis. *J. Affect. Disord.* **2024**, *356*, 459–469. [CrossRef] [PubMed]
35. Coghlan, S.; Leins, K.; Sheldrick, S.; Cheong, M.; Gooding, P.; D'Alfonso, S. To chat or bot to chat: Ethical issues with using chatbots in mental health. *Digit. Health* **2023**, *9*, 20552076231183542. [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.

computers

MDPI

Article

A Case Study of Computational Thinking Analysis Using SOLO Taxonomy in Scientific–Mathematical Learning

Alejandro De la Hoz Serrano *, Andrés Álvarez-Murillo *, Eladio José Fernández Torrado, Miguel Ángel González Maestre and Lina Viviana Melo Niño

Department of Experimental Science and Mathematics Teaching Area, University of Extremadura, 06006 Badajoz, Spain; eladioft@gmail.com (E.J.F.T.); maestre@unex.es (M.Á.G.M.); lvmelo@unex.es (L.V.M.N.)
* Correspondence: alexdlhoz@unex.es (A.D.l.H.S.); andalvarez@unex.es (A.Á.-M.)

Abstract: Education nowadays requires a certain variety of resources that allow for the acquisition of 21st-century skills, including computational thinking. Educational robotics emerges as a digital resource that supports the development of these skills in both male and female students across different educational stages. However, it is necessary to investigate in depth evaluations that analyze the acquisition of Computational Thinking skills in pre-service teachers, especially when scientific and mathematical content learning programs are designed. This study aims to analyze Computational Thinking skills using the SOLO taxonomy, with an approach to science and mathematics learning, through an intervention based on programming and Educational Robotics. A quasi-experimental design was used on a total sample of 116 pre-service teachers. The SOLO taxonomy categorization was used to associate each level of the taxonomy with the computational concepts analyzed through a quantitative questionnaire. The taxonomy levels associated with Computational Thinking skills correspond to *uni-structural* and *multi-structural* levels. Males presented better results before the intervention, while subsequently, females presented better levels of Computational Thinking, as well as a greater association with the higher complexity level of learning analyzed. In turn, there was a trend between the levels of the SOLO taxonomy and computational concepts, so that an increase in skill for a concept occurs similarly at both the *uni-structural* level and the *multi-structural* level. The SOLO taxonomy is presented as a proper tool for learning assessment since it allows for a more detailed understanding of the quality of students' learning. Therefore, the SOLO taxonomy serves as a valuable resource in the evaluation of Computational Thinking skills.

Keywords: computational thinking; educational robotics; science education; mathematics education; SOLO taxonomy

1. Introduction

In the current era, rapid advances in Computer Science (CS), Robotics, and Artificial Intelligence (AI) are transforming society at an unprecedented pace, creating new demands for educational systems worldwide [1–3]. These technological changes require not only technical knowledge but also the development of transferable problem-solving skills applicable across disciplines. In response, education systems are being challenged to cultivate competencies that go beyond traditional literacy, giving rise to the concept of "code literacy" [1]—a form of digital fluency that equips individuals to participate actively and critically in computationally driven societies.

Among emerging educational priorities, Computational Thinking (CT) has gained significant global attention as a core set of cognitive and procedural skills derived from Computer Science. CT includes abstraction, decomposition, debugging, pattern recognition, and algorithmic thinking [4]. Recognizing its potential to enhance learning across domains, many countries—such as the United States, China, Australia, and members of the European Union—have adopted policies to integrate CT into their educational frameworks, extending its implementation across all educational stages, including higher education [5,6].

However, the incorporation of CT into formal education presents substantial challenges. Chief among them is the limited preparation of teachers, particularly those in pre-service training, to understand and implement CT effectively in classrooms [7]. This gap underscores the importance of equipping future educators with not only the technical tools but also the cognitive frameworks necessary to foster and evaluate CT in students.

In this context, assessing how students learn and apply CT becomes a critical concern. In recent years, various educational taxonomies have been used to assess the development of CT skills. Among them, Bloom's Revised Taxonomy and Webb's Depth of Knowledge (DOK) have been widely applied to describe and measure cognitive complexity [8,9]. However, although useful for categorizing learning objectives, these taxonomies often fall short in assessing how students structure and integrate their knowledge across disciplines and over time. In this context, the Structure of Observed Learning Outcomes (SOLO) taxonomy offers a compelling alternative.

To address this limitation, the Structure of Observed Learning Outcomes (SOLO) taxonomy offers a more dynamic and developmental model for assessing learning. SOLO classifies responses based on the structural complexity of student understanding, ranging from pre-structural (no understanding) to extended abstract (generalization and transfer). This is especially relevant given evidence that CT supports a range of cognitive functions—such as working memory, inhibitory control, and spatial reasoning—that are central to learning in science and mathematics [10–13]. Therefore, evaluating CT through the lens of SOLO allows for a deeper understanding of how learners progress cognitively as they engage in computational problem-solving, rather than simply what they can produce.

This study addresses a key gap in the literature: the absence of robust and developmentally sensitive frameworks for evaluating the quality and depth of learning in CT, particularly within the context of pre-service teacher education. To address this gap, the present study adopts the SOLO taxonomy to assess the development of CT skills during an Educational Robotics intervention within a science–mathematics learning environment. This approach allows for a more nuanced analysis of how students structure their knowledge in interdisciplinary CT tasks. The study also seeks to examine the taxonomy's potential to inform the design of instructional strategies and assessment tools that are better aligned with the cognitive and structural demands of CT learning. To this end, the study addresses the following research questions (RQs):

- RQ1: To what extent does the SOLO taxonomy effectively reflect the development of computational thinking skills in pre-service elementary teachers after engagement in an educational robotics program focused on science and mathematics?
- RQ2: Are there gender differences in the SOLO taxonomy levels achieved by pre-service elementary teachers after completing an intervention with educational robotics focusing on developing computational thinking skills in science and mathematics?

To address these questions, the present study builds on a multidisciplinary body of research that explores the role of Computational Thinking in science and mathematics education. In the following section, we outline the theoretical foundations that support the integration of CT into interdisciplinary learning contexts and justify the use of the SOLO taxonomy as a developmental framework for evaluating students' conceptual progression.

2. Theoretical Background

Computational Thinking (CT) has been widely recognized as a core competency for 21st-century learners [14,15] and is explicitly identified as a foundational practice in the Next Generation Science Standards (NGSS) [16]. Originally rooted in the fields of computer science and programming, CT has gradually evolved into a cross-disciplinary construct, with increasing attention directed toward its integration into science, mathematics, technology, and engineering education [13,17,18]. This interdisciplinary expansion is driven in part by the accelerating pace of technological advancement in these fields [19]. Innovations in areas such as bioinformatics, computational statistics, chemometrics, and data analytics have highlighted the need for a computational approach in both scientific investigation and problem-solving across academic and professional contexts [13].

Within science education, CT has emerged as a central focus for both pedagogical development and empirical investigation [20]. Science and computer science share methodological affinities in their use of the scientific method and the engineering design process—both of which involve iterative cycles of hypothesis generation, modeling, refinement, and evaluation. Moreover, both domains depend on core computational constructs such as abstraction, modularity, and algorithmic reasoning to represent and understand complex systems [20]. A similar convergence has been noted in the relationship between CT and mathematics, where CT supports mathematical learning through shared cognitive mechanisms such as problem decomposition, pattern recognition, and symbolic representation [21–23]. However, as Lockwood and De Chenne [24] note, not all mathematical concepts translate seamlessly into programming environments, and students often face significant cognitive dissonance when navigating between computational and mathematical logics [18]. Nonetheless, scholars such as Sinclair and Patterson [25] maintain that CT and mathematics remain deeply intertwined and highlight the opportunity for pedagogical innovation through their integration.

Despite growing interest in CT integration, significant gaps persist in the literature, particularly in the development of robust and context-sensitive assessment models. Shumway et al. [26] and Ye et al. [27] both underscore the lack of systematic reviews and empirical studies that explore CT's intersection with mathematics education, particularly within interdisciplinary or project-based learning environments. Furthermore, there remains little consensus on how to meaningfully assess the depth and structure of students' CT development, especially when embedded in cross-disciplinary settings such as science–mathematics curricula. While existing studies report promising outcomes from integrative interventions [28–31], they often rely on generalized evaluation frameworks or competency rubrics that emphasize surface-level indicators such as correctness or procedural fluency.

To address these limitations, Weintrop et al. [13] proposed a taxonomy tailored specifically to CT in mathematics and science education, comprising four core domains: data practices, modeling and simulation practices, computational problem-solving practices, and systems thinking practices (see Figure 1). This contribution has advanced the field by providing a domain-specific framework for instructional design. However, the assessment strategies aligned with such taxonomies often remain generic or summative in nature, failing to account for the evolving complexity of students' conceptual understanding throughout the learning process.

However, despite the growing body of evidence on the positive effects of CT-related interventions, a critical gap persists in the literature regarding how to assess the depth and structure of students' learning in CT. Most traditional assessment methods—such as performance rubrics or standardized tests based on taxonomies like Bloom's—tend to emphasize accuracy or hierarchical levels of cognitive achievement [8,9].

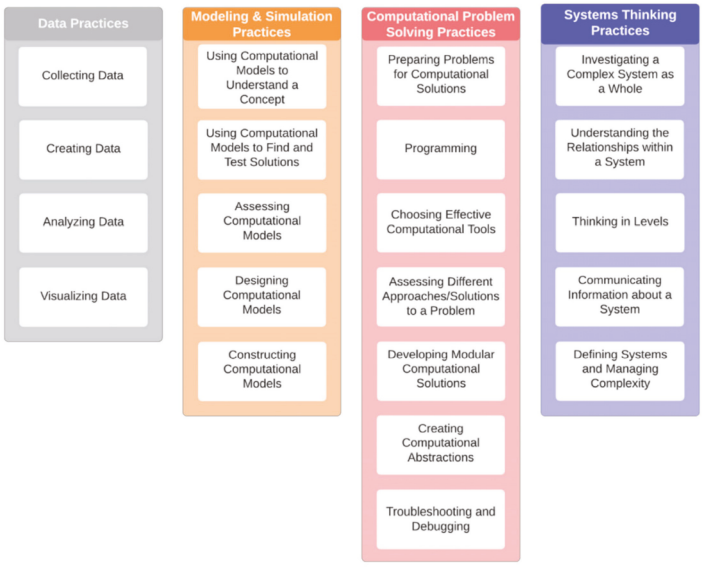


Figure 1. Taxonomy of Computational Thinking in mathematics and science [10].

To address this limitation, the present study adopts the Structure of Observed Learning Outcomes (SOLO) taxonomy as a framework for evaluating the quality of CT learning. Unlike other taxonomies, SOLO focuses on the structural complexity of students' thinking, ranging from *uni-structural* responses to more relational and abstract levels of understanding.

In this context, the SOLO taxonomy emerges as a promising yet underutilized alternative for assessing CT development. Unlike Bloom's taxonomy, which emphasizes types of cognitive processes (e.g., remembering, evaluating), or Webb's Depth of Knowledge (DoK), which focuses on task complexity, SOLO centers on the structural transformation of learners' responses—from fragmented to integrated and abstract understandings [32]. This feature is particularly relevant for CT, where learning involves not only acquiring isolated skills but synthesizing them into transferable strategies across domains. The SOLO taxonomy captures both quantitative and qualitative dimensions of learning progression (see Figure 2), making it well suited to interdisciplinary contexts where CT serves as a bridge between conceptual domains.

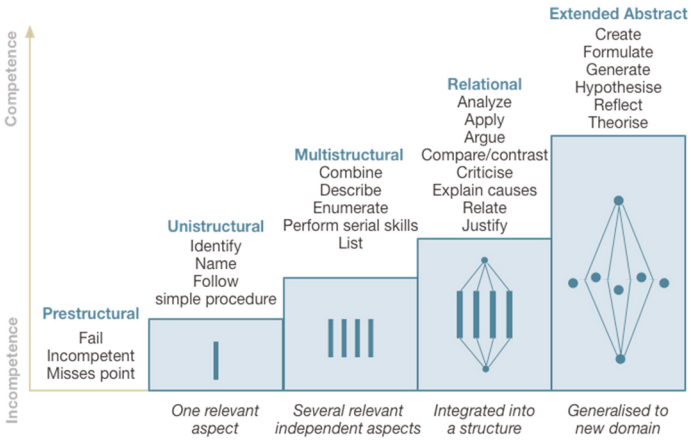


Figure 2. Categorization of the SOLO taxonomy [32].

Furthermore, SOLO provides a non-hierarchical model that allows researchers to track students' movement across five levels of cognitive complexity: *pre-structural*, *uni-structural*, *multi-structural*, *relational*, and *extended abstract*. These levels reflect not only the accumulation of content but also the increasing integration and generalization of

knowledge, a key attribute of CT expertise [33]. Table 1 illustrates how specific action verbs align with each level of taxonomy, offering educators and researchers a practical tool for designing assessments that correspond to targeted learning outcomes [34].

Table 1. Verbs and terms associated with the SOLO taxonomy levels [34].

Level 2 *Uni-Structural*	**Level 3** *Multi-Structural*	**Level 4** *Relational*	**Level 5** *Extended Abstract*
Paraphrase	Combine	Analyze	Theorize
Define	Classify	Compare	Generalize
Identify	Structure	Contrast	Hypothesize
Count	Describe	Integrate	Predict
Name	Enumerate	Relate	Judge
Recite	List	Explain causes	Reflect
Follow Instructions	Algorithmic approach	Apply theory	Transfer theory

While prior studies have applied the SOLO taxonomy to evaluate learning outcomes in fields such as computer science and statistics [35–37], its application to CT assessment in interdisciplinary STEM contexts remains limited. Several studies [38–40] advocate for its application in the teaching of various disciplines such as science and mathematics, with the aim of enhancing the assessment of learning quality at any educational level. The use of this tool can offer a more nuanced understanding of how students learn subject content, thereby supporting the design of student-centered interventions and improving the overall teaching–learning process

Some researchers have explored its utility in programming education and found it effective in capturing the nuanced development of computational proficiency beyond traditional rubrics. For example, Sari et al. [41] used SOLO to explore gender-based differences in mathematics learning, finding male students exhibited more frequent extended abstract responses. Similarly, Tsan et al. [42] observed parallel trends in programming contexts, while Obreque et al. [43] reported higher performance among female students in scientific content tasks using SOLO analysis.

Furthermore, as evidenced in previous studies [8,9], there is a growing interest in assessing Computational Thinking (CT) beyond the confines of computer science, particularly within teacher education programs, where future educators are required to develop transferable problem-solving competencies. These studies underscore the necessity for more refined assessment instruments, although they primarily rely on generalized evaluation frameworks or competency rubrics, rather than employing structured taxonomies such as SOLO. This gap presents an opportunity for the broader integration of the SOLO taxonomy to more effectively capture the qualitative depth of CT development within these emerging educational contexts.

Despite these promising insights, the current body of research remains fragmented and limited in scope, particularly when it comes to applying the SOLO taxonomy within interdisciplinary learning environments that combine science, mathematics, and Computational Thinking. Most studies focus either on computing-specific contexts or general subject domains, without addressing the unique cognitive demands that arise when CT is embedded into non-computing curricular areas through innovative methodologies such as Educational Robotics.

3. Materials and Methods

3.1. Study Design and Sample

This study employed a pre-experimental design with a quantitative approach, using both descriptive and inferential statistical techniques. The sample consisted of 116 pre-

service teachers enrolled in the third year of a four-year Primary Education degree program at a Spanish university.

Of the total participants, 38 were male and 78 were female. This gender ratio is consistent with the typical demographic profile of Primary Education degree programs in Spain, where female enrollment significantly outnumbers male enrollment. This group was chosen specifically because they had not received formal instruction in Computational Thinking (CT) or Educational Robotics prior to the intervention. Their academic stage and profile make them an ideal target population for examining the development of CT-related skills in future educators of science and mathematics.

3.2. Intervention

In order to explore Computational Thinking (CT) skills, a block-based programming training program was implemented. This program consisted of five sessions grounded in constructivist learning approaches, enabling students to develop various skills throughout the proposed challenges, such as creativity, problem-solving, and self-regulation. During these sessions, students explored the wide range of possibilities offered by programming tools such as Scratch 3.0® and the Mind Designer® Educational Robotics Kit, both of which were integrated into teaching and learning contexts involving scientific and mathematical content.

All intervention activities were designed with the aim of fostering fundamental computational concepts such as addresses, loops, conditionals, and functions or sequences. Accordingly, both teacher demonstrations and student work were required to incorporate the use of programming code and blocks aligned with these computational concepts.

The first two sessions were dedicated to the introduction and use of Scratch 3.0® software within a teaching–learning framework focused on mathematical content. The primary goal of these sessions was to employ programming to create mandalas composed of various geometric shapes, which had to be drawn both separately and in overlapping arrangements, using the software's functionalities such as loops, variable creation, and conditionals (see Figure 3).

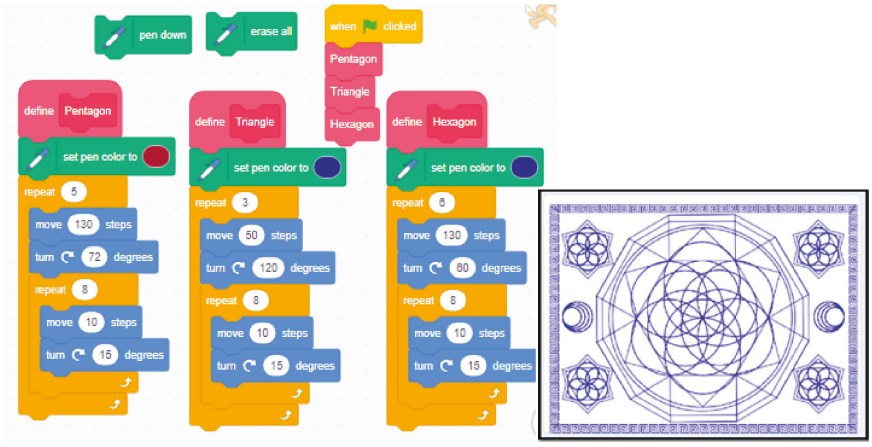


Figure 3. Examples of programming and mandala.

Session 3 consisted of a theoretical explanation and experimentation with the possibilities offered by Educational Robotics (ER) and the Mind Designer® robotics kit. Students had the opportunity to experiment with this resource by completing various programming tasks, similar to those previously explained. The session concluded with an example of activities designed for teaching scientific and mathematical content. In the last two sessions, students, working in cooperative groups of four, were required to create a robotic board as a teaching tool for scientific learning. Specifically, the content focused on the scientific topic

of healthy hydration habits, which allowed for interdisciplinary work with both science and mathematics content.

To measure Computational Thinking, a modified version of the test designed and validated by Román-González [44] was used. The shortened online questionnaire was administered in a similar manner both before and after the intervention. This design is characterized by evaluating different computational concepts aligned with the learning standards set by the 'Computer Science Teachers Association' (CSTA) for computer science education, and it allows for adaptation to different educational stages, as demonstrated in the present study [45]. The computational concepts employed are as follows: *addresses*: instructions that allow for moving the object in different directions; *loops*: this refers to performing the same sequence of instructions multiple times; *conditionals*: this involves understanding the implications of decision-making when carrying out a task or activity; and *functions*: common code modules are typically referred to as functions; they are like labeled drawers in which data can be stored until a program needs them.

The modifications were made with the goal of reducing the time required to complete the questionnaire while preserving the quality of the evaluation of the various computational concepts. Specifically, these adjustments focused on reducing the number of items (from 28 to 12) while ensuring that at least two questions were retained for each computational concept. This reduction was carried out based on the need to include sufficient elements to assess the participants' level of Computational Thinking (CT), with the internal consistency coefficient calculated at 0.79. Data collection and analysis were conducted in accordance with the principles outlined in the Declaration of Helsinki [46].

3.3. Data Analysis

To analyze the level of Computational Thinking (CT) based on the Román-González questionnaire [44], specific questions from the test were selected and mapped to a particular level of the SOLO taxonomy corresponding to each computational concept. The 12 items in the test were carefully examined, and their statements were matched with the verbs described at each level of the SOLO taxonomy, as outlined by Brabrand and Dahl [34]. The use of the SOLO taxonomy, as shown in Appendix A, enabled the classification of each computational concept (*addresses*, *loops*, *conditionals*, and *functions*) into two levels: *uni-structural* and *multi-structural*.

For the quantitative analysis of the collected data, R-2.14.0 software [47] was employed to perform both descriptive and inferential statistics. Given the non-normality of the data ($p < 0.05$), non-parametric statistical tests were applied. Specifically, the Wilcoxon test was used to assess statistically significant differences between pre- and post-intervention results, while the Mann–Whitney U test was utilized to examine differences between the two independent groups, namely male and female students.

4. Results

The results analyzed in this study are presented below. First, the descriptive analysis results are shown, followed by the results of the inferential tests. As observed in Table 2 and Figure 4, regarding the pretest, there is a higher percentage overall in the *uni-structural* dimension (64.14%) compared to the *multi-structural* dimension (53.42%). Specifically, in four of the six computational concepts (*addresses*, *loops*, *conditional while*, and *functions*), a higher percentage is found in the *uni-structural* dimension.

Table 2. Percentage of correct answers in pretest based on gender and SOLO taxonomy dimensions.

Computational Concept		SOLO Dimension	Female	Male	Total
Addresses		*Uni-structural*	83.54%	100%	88.39%
		Multi-structural	68.35%	60.61%	66.07%
Loops	Repeat	*Uni-structural*	75.95%	84.85%	78.57%
		Multi-structural	35.44%	45.45%	38.39%
Conditional	Simple	*Uni-structural*	30.38%	48.48%	35.71%
		Multi-structural	64.56%	81.82%	69.64%
	Compound	*Uni-structural*	63.29%	63.64%	63.39%
		Multi-structural	73.42%	81.82%	75.89%
	While	*Uni-structural*	45.57%	63.64%	50.89%
		Multi-structural	29.11%	18.18%	25.89%
Functions		*Uni-structural*	64.56%	75.76%	67.86%
		Multi-structural	36.71%	63.64%	44.64%
Total		*Uni-structural*	60.55%	72.73%	64.14%
		Multi-structural	51.27%	58.59%	53.42%

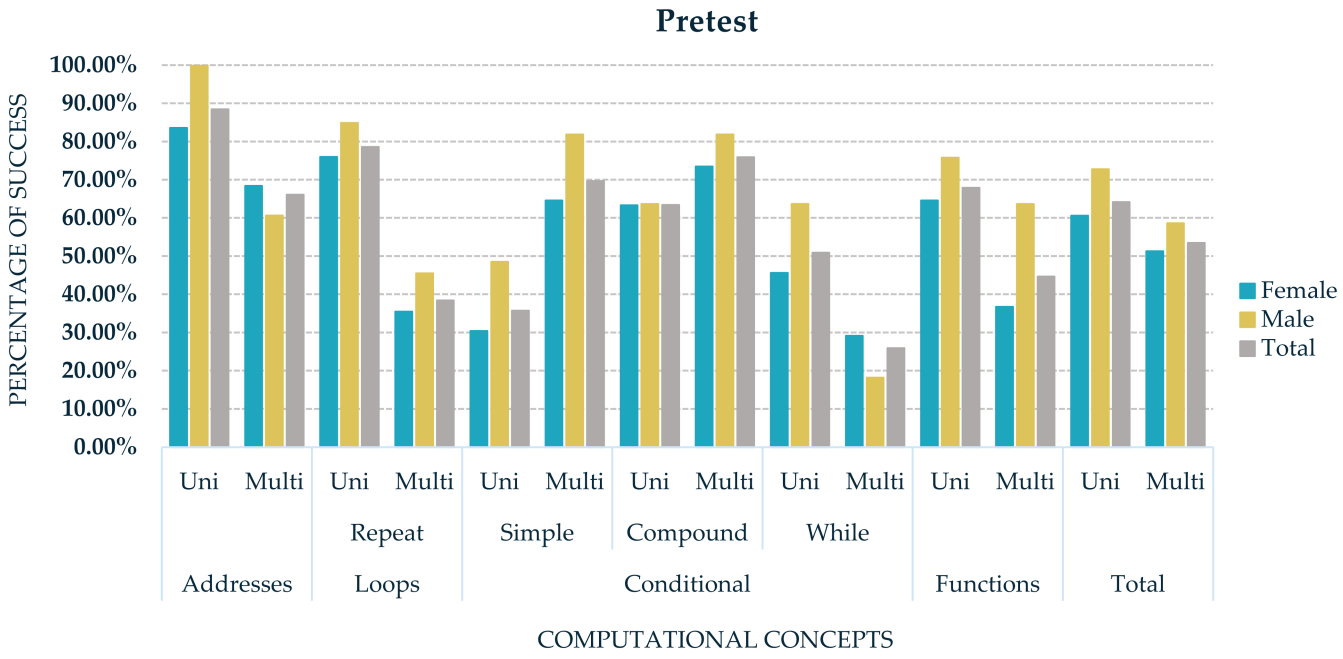


Figure 4. Percentage of correct answers in pretest based on gender and SOLO taxonomy dimensions.

Focusing on the sex of the study sample, both groups show a higher percentage in the *uni-structural* dimension for the same computational concepts as observed overall. When comparing the two groups, the male group has a higher percentage of correct answers across all six computational concepts in the *uni-structural* dimension. In contrast, in the *multi-structural* dimension, the female group shows a higher percentage in two of the six concepts (*addresses* and *while conditionals*). Overall, the male group exhibits a higher percentage than the female group in both dimensions.

Table 3 and Figure 5 display the percentage of correct responses in the post-test. It can be observed that, overall, there is a higher percentage of correct answers in the *uni-structural* dimension (73.21%) compared to the *multi-structural* dimension (63.84%). Specifically, in

four out of the six computational concepts (*addresses*, *loops*, *while conditionals*, and *functions*), a higher percentage is found in the *uni-structural* dimension.

Table 3. Percentage of correct answers in post-test based on gender and SOLO taxonomy dimensions.

Computational Concept		SOLO Dimension	Female	Male	Total
Addresses		*Uni-structural*	89.87%	90.91%	90.18%
		Multi-structural	77.22%	72.73%	75.89%
Loops	Repeat	*Uni-structural*	78.48%	84.85%	80.36%
		Multi-structural	50.63%	48.48%	50.00%
Conditional	Simple	*Uni-structural*	37.97%	48.48%	41.07%
		Multi-structural	73.42%	72.73%	73.21%
	Compound	*Uni-structural*	63.29%	60.61%	62.50%
		Multi-structural	74.68%	75.76%	75.00%
	While	*Uni-structural*	82.28%	54.55%	74.11%
		Multi-structural	43.04%	39.39%	41.96%
Functions		*Uni-structural*	89.87%	93.94%	91.07%
		Multi-structural	67.09%	66.67%	66.96%
Total		*Uni-structural*	73.63%	72.22%	73.21%
		Multi-structural	64.35%	62.63%	63.84%

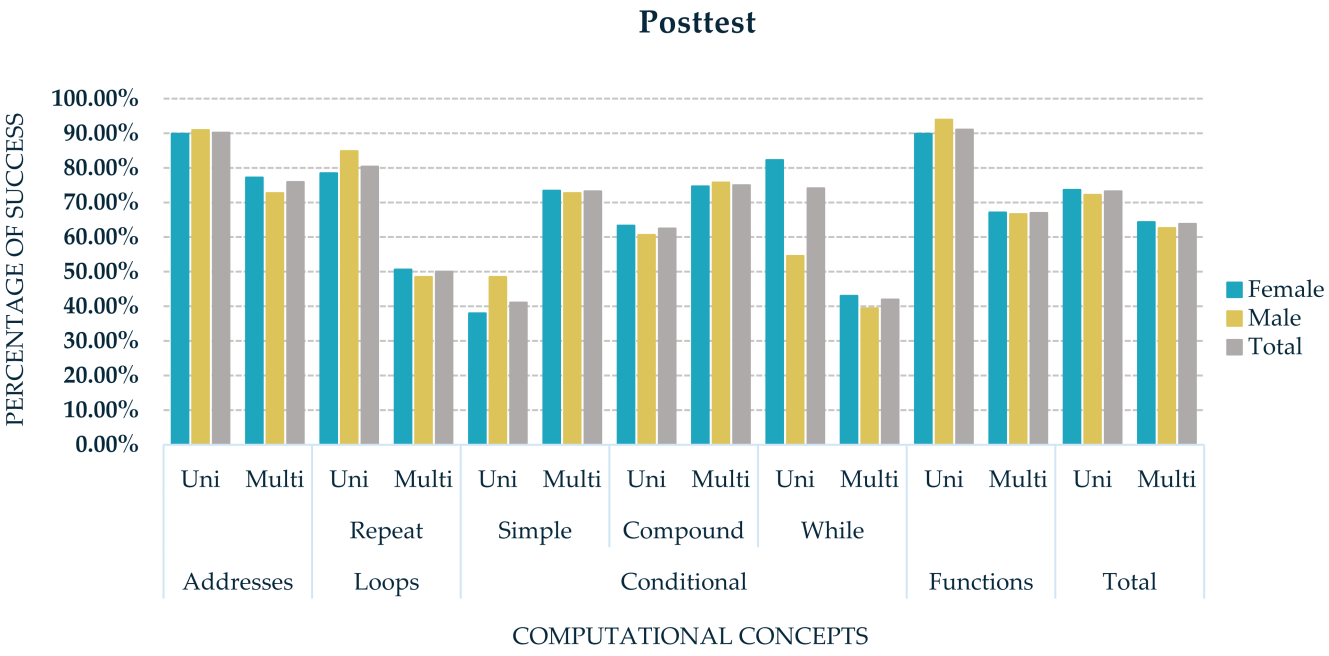


Figure 5. Percentage of correct answers in post-test based on gender and SOLO taxonomy dimensions.

When considering the sex of the study sample, both groups show a higher percentage in the *uni-structural* dimension for the same computational concepts as observed in the overall results. Regarding the differences between the two, the male group exhibits a higher percentage in four out of the six computational concepts (*addresses*, *loops*, *simple conditionals*, and *functions*) in the *uni-structural* dimension. In contrast, in the *multi-structural* dimension, the female group shows a higher percentage across all concepts, except for *compound conditionals*. Overall, in both dimensions, the female group presents a higher percentage.

The results from the Mann–Whitney U test (Table 4) show that statistically significant differences ($p < 0.05$) between female and male participants were observed. Although statistically significant differences were found between male and female participants in certain dimensions and computational concepts, these differences are not consistent across all variables measured.

Table 4. Mann–Whitney U test results for pretest and post-test according to SOLO taxonomy dimensions between male and female students.

Computational Concept		SOLO Dimension	Mann–Whitney U Test			
			Pretest		Posttest	
			S	*p*	S	*p*
Addresses		*Uni-structural*	1089	0.01 *	1290	0.87
		Multi-structural	1203	0.43	1245	0.62
Loops	Repeat	*Uni-structural*	1188	0.30	1221	0.44
		Multi-structural	1173	0.33	1276	0.84
Conditional	Simple	*Uni-structural*	1068	0.07	1167	0.31
		Multi-structural	1079	0.07	1295	0.94
	Compound	*Uni-structural*	1299	0.98	1269	0.79
		Multi-structural	1194	0.35	1290	0.91
	While	*Uni-structural*	1068	0.08	942	0.00 *
		Multi-structural	1161	0.23	1256	0.73
Functions		*Uni-structural*	1158	0.25	1251	0.50
		Multi-structural	953	0.01 *	1298	0.97
Total		*Uni-structural*	924	0.01 *	1240	0.68
		Multi-structural	1068	0.13	1289	0.93

* Statistically significant differences ($p < 0.05$).

In relation to the pretest, statistically significant differences were found in the concepts of *addresses* (*uni-structural* dimension) and *functions* (both the *uni-structural* and *multi-structural* dimensions), as well as in the overall *uni-structural* dimension. Regarding the post-test, statistically significant differences were observed in the concept of *while conditionals* (*uni-structural* dimension). In both pretest cases, male participants showed a higher percentage of correct answers, whereas in the posttest, it was the female participants who demonstrated a higher percentage.

Furthermore, for the concepts of *simple conditionals* (both the *uni-structural and multi-structural* dimensions) and *compound conditionals* (*uni-structural* dimension), there are indications of statistical significance ($p \approx 0.05$).

Regarding the differences between the pretest and post-test results, Table 5 indicates statistically significant ($p < 0.05$) improvements overall, largely driven by the female participants, who exhibited significant gains across multiple concepts and SOLO dimensions. In contrast, male participants showed limited improvement, with significant differences observed only for one concept (*while conditionals*, *multi-structural* dimension).

Table 5. Wilcoxon test by gender and SOLO taxonomy dimensions between pretest and post-test.

Computational Concept		SOLO Dimension	Wilcoxon Test					
			M		F		T	
			S	*p*	S	*p*	S	*p*
Addresses		*Uni-structural*	6	0.15	88	0.28	137	0.70
		Multi-structural	37	0.30	252	0.24	475	0.12
Loops	Repeat	*Uni-structural*	18	1.00	217	0.72	351	0.75
		Multi-structural	42	0.81	222	0.04 *	450	0.06
Conditional	Simple	*Uni-structural*	39	1.00	312	0.34	561	0.40
		Multi-structural	30	0.35	285	0.25	493	0.56
	Compound	*Uni-structural*	35	0.43	259	0.12	475	0.04 *
		Multi-structural	30	0.80	155	0.07	315	0.06
	While	*Uni-structural*	72	0.46	76	0.00 *	344	0.00 *
		Multi-structural	5.0	0.02 *	315	0.09	408	0.00 *
Functions		*Uni-structural*	11	0.07	77.5	0.00 *	143	0.00 *
		Multi-structural	56	0.82	136	0.00 *	378	0.00 *
Total		*Uni-structural*	32	0.91	78	0.00 *	111	0.00 *
		Multi-structural	32	0.53	78	0.00 *	111	0.01 *

* Statistically significant differences ($p < 0.05$).

Statistically significant differences ($p < 0.05$) were observed in the post-test for the concepts *while conditionals* and *functions* (in both the *uni-structural* and *multi-structural* dimensions), as well as in the overall scores, and for the *compound conditionals (uni-structural* dimension). Additionally, near-significant values were detected for *loops* (*multi-structural* dimension) and *compound conditionals* (*multi-structural* dimension).

With regard to gender-based differences, statistically significant improvements were identified for the concept *while conditionals* (*multi-structural* dimension) for male participants. Among female participants, statistically significant differences were found for the concepts *loops* (*multi-structural* dimension), *while conditionals* (*uni-structural* dimension), and *functions* (both *uni-structural* and *multi-structural* dimensions), along with significant overall improvement across both dimensions.

These results highlight the utility of the SOLO taxonomy in providing detailed and differentiated insights into learning quality. The taxonomy enabled the identification of specific levels of cognitive complexity, with the *uni-structural* level emerging as the most frequently achieved. Furthermore, improvement was noted across both *uni-structural* and *multi-structural* levels, suggesting enhanced learning outcomes following the intervention.

5. Discussion

The development of programming and computational thinking (CT) skills is increasingly recognized as a fundamental element of 21st-century education—not only within computer science but also across disciplines such as mathematics and science. This growing recognition is reflected in international curricular frameworks and empirical studies, which emphasize the need to equip future teachers with transferable CT skills that can be applied across educational and real-world contexts [35–37]. The present study examined the evolution of CT skills in prospective primary teachers before and after a robotics-based intervention grounded in science and mathematics content, using the SOLO taxonomy as the primary evaluative framework

Regarding RQ1: *How effectively does the SOLO taxonomy reflect changes in the computational thinking skills of prospective primary teachers as a result of participating in an educational robotics program focused on science and mathematics*? The findings clearly support the utility of the SOLO taxonomy in capturing both quantitative and qualitative changes in participants' CT skills. Pre- and post-intervention analyses revealed noticeable cognitive progression, from predominant *uni-structural* levels toward more frequent *multi-structural* responses, in key CT concepts such as *loops*, *conditional*, and *functions*. This shift suggests not only an increase in the volume of knowledge but also an improvement in the structural complexity with which participants organize and apply CT concepts.

While the statistical results indicate significant differences between pre- and post-intervention scores, particularly among female participants, it is important to interpret these changes as evidence of an evolving understanding of computational concepts rather than isolated score improvements. The progression from *uni-structural* to *multi-structural* responses reflects not only increased familiarity with coding elements but also a deeper integration of Computational Thinking processes. This development aligns with the structured nature of the intervention, which combined theoretical explanation with practical problem-solving tasks. Thus, the observed learning gains can be interpreted as a consequence of both cognitive engagement and instructional design, suggesting that targeted interventions can foster meaningful conceptual growth.

These results align with prior research demonstrating the positive impact of robotics-based interventions on CT skill development among future educators [48–50], particularly within disciplinary contexts such as science and mathematics education [51]. More specifically, the current study found that *uni-structural* reasoning remained dominant in both pre- and post-test responses, although an increase in *multi-structural* responses was observed after the intervention. This pattern mirrors findings from a study conducted by Uzumcu and Bay [40], which applied the SOLO taxonomy to CT skill tests before and after a similar educational robotics intervention. In their study, students predominantly displayed *pre-structural* levels prior to the intervention but reached higher levels—including *relational* and *extended abstract* understanding—after participating in the program.

The present study aligns with this trend. Specifically, dominance in the *uni-structural* level was observed for the same concepts (*addresses*, *loops*, *while conditionals*, and *functions*) before and after the intervention. A clear trend is evident between the two levels of the SOLO taxonomy and the various computational concepts analyzed. Notably, in concepts such as *addresses*, *compound conditionals*, and *functions*, where an increase in correct answers was observed after the intervention, this improvement is reflected similarly at both the *uni-structural* and *multi-structural* levels. Conversely, in concepts like *loops*, *simple conditionals*, and *while conditionals*, where a lower percentage of correct answers was observed post-intervention, the pattern was similar across both levels of the SOLO taxonomy.

The divergence between the present study and that of Uzumcu and Bay [40] lies in the post-intervention depth of cognitive development: while our participants showed improved performance, the majority did not achieve relational or extended abstract levels. This difference may be attributed to several contextual factors, such as the duration and intensity of the intervention, the cognitive demands for integrating CT with scientific and mathematical content, or the participants' prior knowledge and teaching experience. This indicates the need for sustained and scaffolded interventions that allow learners to internalize and apply CT concepts in increasingly complex and interconnected ways.

In parallel, findings from other studies reinforce this trend. For instance, Fojcik et al. [36] investigated whether introductory programming courses in non-CS domains yield substantial learning gains, while Karvounidis et al. [37] and Almeida et al. [35] reported high frequencies of *uni-structural* and *multi-structural* reasoning in students' programming work

across languages. Notably, these studies also observed a transitional relationship between *multi-structural* and *relational* levels, reinforcing the need for further research to explore how students advance beyond surface-level CT understanding [40].

The SOLO taxonomy proves to be an effective tool for evaluating coding skills and providing guidance for structuring teaching strategies through the creation of personalized teaching pathways [46]. By identifying the SOLO level at which learners operate for each concept, educators can design personalized instructional pathways that close knowledge gaps and foster progression toward higher-order thinking. This perspective is supported by research in science and mathematics education, which similarly reports the predominance of surface-level SOLO responses and advocates for explicitly designed interventions that foster relational and extended abstract understanding [38,39,52,53].

Addressing RQ2: *What differences, if any, exist between male and female prospective primary teachers in the SOLO taxonomy levels attained after completing a robotics-based intervention designed to enhance computational thinking within science and mathematics education*? The data reveal statistically significant differences between male and female participants, both pre- and post-intervention.

Initially, male participants displayed higher CT proficiency, a pattern corroborated by previous studies [41,42,51,54]. However, following the intervention, female participants matched or surpassed their male peers, particularly in the *multi-structural* dimension, indicating a marked improvement in their ability to engage with complex CT constructs. These findings align with those of Obreque et al. [43] and De la Hoz et al. [51], suggesting that well-designed, inclusive interventions can play a pivotal role in reducing gender disparities in CT education.

Similarly, the results show that female participants exhibited statistically significant differences in more computational concepts and SOLO taxonomy levels than their male counterparts, suggesting a higher level of learning achieved by the female participants. While female participants demonstrated significant improvements across a wider range of computational concepts and SOLO levels, male participants showed more limited progress. This may be due to higher initial baseline performance among male participants, suggesting possible prior exposure or experience with computational thinking. Alternatively, the instructional design of the intervention may have better aligned with the learning styles or engagement patterns of the female participants. Given the 2:1 female-to-male ratio in the sample, these results also raise questions about statistical weighting and generalizability. Future studies should aim for balanced samples and potentially adapt the intervention to ensure it supports learners with varying levels of prior knowledge.

These findings reinforce the value of the SOLO taxonomy as both a diagnostic and formative tool in computational thinking education. Furthermore, they underscore the importance of sustained and contextually grounded interventions—such as those based on scientific and mathematical activities and the use of educational robotics—in fostering core CT competencies among future educators. However, the persistent difficulty in advancing toward the *relational* and *extended abstract* levels highlights the ongoing need for innovation in instructional design. Future research should investigate longitudinal interventions, interdisciplinary applications, and the integration of complementary assessment frameworks to better support the development of deep, transferable CT competencies.

6. Conclusions

This study aimed to examine the effectiveness of an educational robotics intervention—embedded within science and mathematics instruction—in enhancing Computational Thinking (CT) skills among prospective primary school teachers, using the SOLO taxonomy

as a diagnostic and evaluative framework. It also sought to identify whether gender-based differences emerged in the SOLO levels achieved following the intervention.

The findings provide robust evidence of the SOLO taxonomy's capacity to reveal qualitative changes in students' CT competencies, particularly in concepts like *while conditionals* and *functions*, where a notable shift from *uni-structural* to *multi-structural* levels was observed. These results indicate not only an increase in content knowledge but also a deeper structural organization of that knowledge—reflecting meaningful cognitive development.

Gender differences were also identified. While male participants initially demonstrated stronger performance, female participants exhibited greater gains after the intervention, surpassing their male peers in several key CT dimensions, particularly in the *multi-structural* dimension. This suggests that inclusive and contextually grounded pedagogical approaches, such as the use of educational robotics in STEM content areas, can contribute to closing gender gaps in CT learning.

Nonetheless, persistent challenges in advancing participants toward the *relational* and *extended abstract* levels highlight the need for further refinement of instructional strategies. Addressing these higher-order dimensions of thinking will be essential to fostering transferable and sophisticated CT skills.

The SOLO taxonomy provided detailed insights into students' learning, as a formative assessment tool and a guide for the design of personalized, equity-focused instructional practices.

7. Limitations of the Study and Future Lines of Research

This study analyzes Computational Thinking (CT) skills through a questionnaire that has been supported and utilized in multiple recent studies. However, the associative analysis with the SOLO taxonomy levels highlights more complex levels that are not fully assessed. Therefore, it is suggested to expand the questionnaires to better evaluate the learning levels within each stage of the taxonomy, allowing for a more specific determination of students' skills. Additionally, the study recommends examining the students' immediate environment as a relevant factor influencing their learning of these skills. This would provide a more comprehensive understanding of the factors affecting CT skill development and help to design more targeted educational interventions.

Author Contributions: Conceptualization, A.D.l.H.S. and A.Á.-M.; methodology, A.D.l.H.S. and A.Á.-M.; software, A.D.l.H.S.; validation, A.Á.-M., E.J.F.T., M.Á.G.M. and L.V.M.N.; formal analysis, A.D.l.H.S. and A.Á.-M.; investigation, A.D.l.H.S., A.Á.-M., E.J.F.T., M.Á.G.M. and L.V.M.N.; resources, A.D.l.H.S. and A.Á.-M.; data curation, A.D.l.H.S. and A.Á.-M.; writing—original draft preparation, A.D.l.H.S.; writing—review and editing, A.Á.-M., E.J.F.T., M.Á.G.M. and L.V.M.N.; visualization, A.D.l.H.S., A.Á.-M., E.J.F.T., M.Á.G.M. and L.V.M.N.; supervision, A.Á.-M., E.J.F.T., M.Á.G.M. and L.V.M.N.; project administration, L.V.M.N.; funding acquisition, A.D.l.H.S. and L.V.M.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the project PID2020-115214RB-I00 funded by MCIN/AEI/10.13039/50110001103.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki. Ethics Committee Name: approved by the Ethics Committee of the Extremadura University. (139/2023, on 28 September 2023).

Data Availability Statement: The original contributions presented in this study are included in the article.

Acknowledgments: This work was supported by The Ministry of Education and Vocational Training for granting a predoctoral contract (FPU20/04959).

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

Appendix A

Items from Román-González Questionnaire [53] selected for association with the SOLO taxonomy dimensions.

Computational Concept		SOLO Taxonomy Dimension	Items of Román-González Questionnaire [53] Associated with SOLO Taxonomy
Addresses		*Uni-structural*	3
		Multi-structural	4
Loops		*Uni-structural*	7
		Multi-structural	8
Conditional	Simple	*Uni-structural*	16
		Multi-structural	14
	Compound	*Uni-structural*	19
		Multi-structural	18
	While	*Uni-structural*	24
		Multi-structural	22
Functions		*Uni-structural*	28
		Multi-structural	25

References

1. National Institute of Educational Technologies and Teacher Training. School of Computational Thinking and Artificial Intelligence 2021/22. From Teacher Training to Change in Methodology. Research Finding. Ministry of Education and Vocational Training. 2023. Available online: https://code.intef.es/wp-content/uploads/2023/04/09_22_Experimentacion_Investigacion-EPCIA-21-22_Investigacion-R3_ing.pdf (accessed on 13 January 2025).
2. Salas-Pilco, S.Z. The impact of AI and robotics on physical, social-emotional and intellectual learning outcomes: An integrated analytical framework. *Br. J. Educ. Technol.* **2020**, *51*, 1808–1825. [CrossRef]
3. Yim, I.H.Y.; Su, J. Artificial intelligence (AI) learning tools in K-12 education: A scoping review. *J. Comput. Educ.* **2025**, *12*, 93–131. [CrossRef]
4. Román-González, M.; Moreno-León, J.; Robles, G. Complementary Tools for Computational Thinking Assessment. In Proceedings of the International Conference on Computational Thinking Education—CTE 2017 The Education University of Hong Kong, Hong Kong, China, 13–15 July 2017.
5. Grover, S.; Pea, R. Computational thinking in K-12: A review of the state of the field. *Educ. Res.* **2013**, *42*, 38–43. [CrossRef]
6. Bocconi, S.; Chioccariello, A.; Kampylis, P.; Dagienė, V.; Wastiau, P.; Engelhardt, K.; Earp, J.; Stupurienė, G. *Reviewing Computational Thinking in Compulsory Education State of Play and Practices from Computing Education*; Inamorato, A., Cachia, R., Giannoutsou, N., Punie, Y., Eds.; Publications Office of the European Union: Luxembourg, 2022. Available online: https://publications.jrc.ec.europa.eu/repository/handle/JRC128347 (accessed on 28 March 2025).
7. Mukasheva, M.; Omirzakova, A. Computational thinking assessment at primary school in the context of learning programming. *World J. Educ. Technol. Curr. Issues.* **2021**, *13*, 336–353. [CrossRef]
8. Vrachnos, E.; Jimoyiannis, A. Secondary Education Students' Difficulties in Algorithmic Problems with Arrays: An Analysis Using the SOLO Taxonomy. *Themes Sci. Technol. Educ.* **2017**, *10*, 31–52.
9. Yadav, A.; Gretter, S.; Good, J.; McLean, T. Computational thinking in teacher education. In *Emerging Research, Practice, and Policy on Computational Thinking. Educational Communications and Technology: Issues and Innovations*; Rich, P., Hodges, C., Eds.; Springer: Cham, Switzerland, 2017. [CrossRef]
10. Bers, M.; Strawhacker, A.; Sullivan, A. *The State of the Field of Computational Thinking in Early Childhood Education*; Report No. 274; OECD Publishing: Paris, France, 2022. [CrossRef]
11. Nouri, J.; Zhang, L.; Mannila, L.; Norén, E. Development of computational thinking, digital competence and 21st century skills when learning programming in K-9. *Educ. Inq.* **2019**, *11*, 1–17. [CrossRef]

12. Sun, L.; You, X.; Zhou, D. Evaluation and development of STEAM teachers' computational thinking skills: Analysis of multiple influential factors. *Educ. Inf. Technol.* **2023**, *28*, 14493–14527. [CrossRef]
13. Weintrop, D.; Beheshti, E.; Horn, M.; Orton, K.; Jona, K.; Trouille, L.; Wilensky, U. Defining computational thinking for mathematics and science classrooms. *J. Sci. Educ. Technol.* **2016**, *25*, 127–147. [CrossRef]
14. Wing, J.M. Computational thinking. *Commun. ACM* **2016**, *49*, 33–35. [CrossRef]
15. Grover, S.; Pea, R. Computational Thinking: A competency whose time has come. In *Computer Science Education: Perspectives on Teaching and Learning in School*; Sentance, S., Barendsen, E., Carsten, S., Eds.; Bloomsbury Publishing: London, UK, 2018; pp. 19–38.
16. NGSS Lead States. *Next Generation Science Standards: For States, by States*; The National Academies Press: Washington, DC, USA, 2013.
17. Resnick, M.; Rusk, N. Coding at a crossroads. *Commun. ACM* **2020**, *63*, 120–127. [CrossRef]
18. Cui, Z.; Ng, O. The interplay between mathematical and computational thinking in primary 1. school students' mathematical problem-solving within a programming environment. *J. Educ. Comput. Res.* **2021**, *59*, 988–1012. [CrossRef]
19. Bailey, D.H.; Borwein, J.M. High-precision numerical integration: Progress and challenges. *J. Symb. Comput.* **2020**, *46*, 741–754. [CrossRef]
20. Herrero, J.F.Á.; Bautista, C.V. Didáctica de las ciencias, ¿de dónde venimos y hacia dónde vamos? *UTE Teach. Technol. (Univ. Tarracon.)* **2019**, *7*, 5–19. [CrossRef]
21. Pei, C.; Weintrop, D.; Wilensky, U. Cultivating computational thinking practices and mathematical habits of mind in Lattice Land. *Math. Think. Learn.* **2018**, *20*, 75–89. [CrossRef]
22. Gadanidis, G. Coding as a Trojan Horse for mathematics education reform. *J. Comput. Math. Sci. Teach.* **2015**, *34*, 155–173.
23. Rambally, G. Integrating computational thinking in discrete structures. In *Emerging Research, Practice, and Policy on Computational Thinking*; Springer: Cham, Switzerland, 2017; pp. 99–119.
24. Lockwood, E.; De Chenne, A. Enriching students' combinatorial reasoning through the use of loops and conditional statements in Python. *Int. J. Res. Undergrad. Math. Educ.* **2022**, *6*, 303–346. [CrossRef]
25. Sinclair, N.; Patterson, M. The dynamic geometrisation of computer programming. *Math. Think. Learn.* **2018**, *20*, 54–74. [CrossRef]
26. Shumway, J.F.; Welch, L.E.; Kozlowski, J.S.; Clarke-Midura, J.; Lee, V.R. Kindergarten students' mathematics knowledge at work: The mathematics for programming robot toys. *Math. Think. Learn.* **2023**, *25*, 380–408. [CrossRef]
27. Ye, H.; Liang, B.; Ng, O.L.; Chai, C.S. Integration of computational thinking in K-12 mathematics education: A systematic review on CT-based mathematics instruction and student learning. *Int. J. STEM Educ.* **2023**, *10*, 3. [CrossRef]
28. Chevalier, M.; Giang, C.; Piatti, A.; Mondada, F. Fostering computational thinking through educational robotics: A model for creative computational problem solving. *Int. J. STEM Educ. Res.* **2020**, *7*, 39. [CrossRef]
29. De la Hoz, A.; Melo, L.; Álvarez, A.; Cañada, F.; Cubero, J. The Promotion of Healthy Hydration Habits through Educational Robotics in University Students. *Healthcare* **2023**, *11*, 2160. [CrossRef]
30. De la Hoz, A.; Cañada, F.; Melo, L.V.; Alvarez, A.; Cubero, J. Teaching Human Hydration Science Content Through Computational Thinking and Educational Robotics. In Proceedings of the 15th Conference of the European Science Education Research Association (ESERA), Cappadocia, Turkey, 28 August–1 September 2023.
31. Jaipal-Jamani, K.; Angeli, C. Effect of robotics on elementary preservice teachers' self- efficacy, science learning, and computational thinking. *J. Sci. Educ. Technol.* **2017**, *26*, 175–192. [CrossRef]
32. Biggs, J.B.; Collis, K.F. *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)*; Academic Press: Cambridge, MA, USA, 1982.
33. Biggs, J. What the student does: Teaching for enhanced learning. *High. Educ. Res. Dev.* **2012**, *31*, 39–55. [CrossRef]
34. Brabrand, C.; Dahl, B. Using the SOLO taxonomy to analyze competence progression of university science curricula. *High. Educ.* **2009**, *58*, 531–549. [CrossRef]
35. Almeida, A.; Araújo, E.; Figueiredo, J. Avaliando a Construção do Conhecimento em Programação Através da Taxonomia SOLO. In Proceedings of the Anais do XXXI Simpósio Brasileiro de Informática na Educação, Natal, Brazil, 24–28 November 2020; pp. 1813–1822.
36. Fojcik, M.K.; Fojcik, M.; Sande, O.; Refvik, K.A.; Frantsen, T.; Bye, H. A content analysis of SOLO-levels in different computer programming courses in higher education. In Proceedings of the Norsk IKT-Konferanse for Forskning og Utdanning, Stavanger, Norway, 24–25 November 2020.
37. Karvounidis, T.; Ladias, A.; Douligeris, C. Assessment of data carriers with the SOLO taxonomy in Scratch. In Proceedings of the 7th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Ioannina, Greece, 23–25 September 2022; pp. 1–7.
38. Vidal-Szabó, P.; Estrella, S. Conocimiento Estadístico Especializado en Profesores de Educación Básica, basado en la taxonomía SOLO. *Rev. Chil. Educ. Matemática* **2021**, *13*, 134–148. [CrossRef]
39. Obreque, A.S.; Uribe, M.E.M.; Richards, G.R. El aprendizaje de la Biología desde la taxonomía SOLO: Niveles SOLO en estudiantes universitarios. *Contextos Estud. Humanidades Cienc. Soc.* **2003**, *10*, 199–212.

40. Uzumcu, O.; Bay, E. The effect of computational thinking skill program design developed according to interest driven creator theory on prospective teachers. *Educ. Inf. Technol.* **2021**, *26*, 565–583. [CrossRef]
41. Sari, D.I.; Budayasa, I.K.; Juniati, D. The analysis of probability task completion; Taxonomy of probabilistic thinking-based across gender in elementary school students. In *AIP Conference Proceedings*; AIP Publishing: New York, NY, USA, 2017; Volume 1868.
42. Tsan, J.; Boyer, K.E.; Lynch, C.F. How early does the CS gender gap emerge? A study of collaborative problem solving in 5th grade computer science. In Proceedings of the 47th ACM Technical Symposium on Computing Science Education, Memphis, TN, USA, 2–5 March 2016; pp. 388–393.
43. Obreque, A.S.; Salvatierra, M.O.; Danilo, D.L. Evaluation of school performance in students with excellent marks in biology. *Rev. Espac.* **2018**, *39*, 31.
44. Román-González, M. Codigoalfabetización y Pensamiento Computacional en Educación Primaria y Secundaria: Validación de un Instrumento y Evaluación de Programas. Ph.D. Thesis, National University of Distance Education, Madrid, Spain, 2016.
45. Lafuente Martínez, M.; Lévêque, O.; Benítez, I.; Hardebolle, C.; Zufferey, J.D. Assessing Computational Thinking: Development and Validation of the Algorithmic Thinking Test for Adults. *J. Educ. Comput. Res.* **2022**, *60*, 1436–1463. [CrossRef]
46. World Medical Association. Declaration of Helsinki—Ethical Principles for Medical Research Involving Human Subjects. 2022. Available online: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ (accessed on 17 October 2024).
47. Ocaña, R. *Discovering R-Commander*, 3rd ed.; Andalusian School of Public Health (EASP): Andalusia, Spain, 2019. Available online: http://www.repositoriosalud.es/handle/10668/2574 (accessed on 28 May 2024).
48. Yun, M.; Crippen, K.J. Computational thinking integration into pre-service science teacher education: A systematic review. *J. Sci. Teach. Educ.* **2025**, *36*, 225–254. [CrossRef]
49. Nannim, F.A.; Ibezim, N.E.; Oguguo, B.C.; Nwangwu, E.C. Effect of project-based Arduino robot application on trainee teachers computational thinking in robotics programming course. *Educ. Inf. Technol.* **2024**, *29*, 13155–13170. [CrossRef]
50. Rodrigues, R.N.; Brito-Costa, S.; Abbasi, M.; Costa, C.; Martins, F. Pre-service teachers' competencies to develop computational thinking: A Portuguese tool to analyse Computational Thinking. *EURASIA J. Math. Sci. Technol. Educ.* **2024**, *20*, em2528. [CrossRef] [PubMed]
51. De la Hoz Serrano, A.; Melo Niño, L.V.; Álvarez Murillo, A.; Martín Tardío, M.Á.; Cañada Cañada, F.; Cubero Juánez, J. Analysis of Gender Issues in Computational Thinking Approach in Science and Mathematics Learning in Higher Education. *Eur. J. Investig. Health Psychol. Educ.* **2024**, *14*, 2865–2882. [CrossRef] [PubMed]
52. Watson, J.; Wright, S.; Fitzallen, N.; Kelly, B. Consolidating understanding of variation as part of STEM: Experimenting with plant growth. *Math. Educ. Res. J.* **2023**, *35*, 961–999. [CrossRef]
53. Mol, S.M.; Matos, D.A.S. Uma análise sobre a Taxonomia SOLO: Aplicações na avaliação educacional. *Estud. Avaliação Educ.* **2019**, *30*, 722–747. [CrossRef]
54. Sun, L.; Liu, J. A gender differential analysis of educational robots' effects on primary teachers' computational thinking: Mediating effect of programming attitudes. *Educ. Inf. Technol.* **2024**, *29*, 19753–19782. [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.

computers

MDPI

Article

Leveraging Technology to Break Barriers in Public Health for Students with Intellectual Disabilities

Georgia Iatraki [1,*], Tassos A. Mikropoulos [1,*], Panos Mallidis-Malessas [1] and Carolina Santos [2]

[1] Department of Primary Education, University of Ioannina, 145110 Ioannina, Greece; panos.mallidis@gmail.com
[2] NOVA National School of Public Health, Public Health Research Centre, Comprehensive Health Research Center, CHRC, NOVA University Lisbon, 1600-560 Lisbon, Portugal; c.santos@ensp.unl.pt
* Correspondence: g.iatraki@uoi.gr (G.I.); amikrop@uoi.gr (T.A.M.)

Abstract: A key goal of inclusive education is to enhance health literacy skills, empowering students with intellectual disabilities (IDs) to access critical information needed to navigate everyday challenges. The COVID-19 pandemic, for example, highlighted unique barriers to preparedness for people with IDs regarding social behavior and decision-making. This study aimed to examine students' awareness and understanding of pandemic outbreaks. Using an inquiry-based approach supported by Digital Learning Objects (DLOs), the research assessed students' knowledge and perceptions of viruses, modes of transmission, and preventive measures. An in-depth visual analysis within a single-subject research design demonstrated that interdisciplinary educational scenarios on infectious diseases can be effective for students with ID, especially when DLOs are integrated with targeted instructional techniques.

Keywords: technology; intellectual disabilities; public health; science literacy

1. Introduction

One of the most prevalent global challenges is the lack of science education, particularly regarding health content, for students with intellectual disabilities (IDs) to cope with a multitude of information on daily life issues and increase their participation in social dialogue [1–3].

During the COVID-19 pandemic outbreak, the ongoing information storm (the so-called "infodemic") created overload and significant barriers in identifying reliable information [4–7]. To overcome disinformation or misinformation, health literacy skills may facilitate navigating across various public health issues, such as infectious diseases [8]. In this direction, people with ID should develop health literacy skills at an adequate level of readiness to foster equally critical thinking, make decisions, and adopt prevention behaviors in periods of epidemic or pandemic outbreaks [9–11].

By engaging students with ID as well as their caregivers in scientific practices, they experience increased service provision [12,13] and acquire science skills to equally participate in emerging challenges in various aspects of daily life [14]. Effective implementation in science classrooms involves activities such as asking questions to solve problems, planning and conducting investigations, analyzing and interpreting data, constructing explanations, using arguments, evaluating, and communicating findings [14,15]. Specifically, inclusive lessons based on inquiry-based learning can enhance comprehension for students with ID, helping them stay focused, reduce memory challenges, and improve attention [16].

Reviews on teaching science to students with ID recommend the connection between basic science content and daily communication using adapted instructional practices in suitable learning environments [17,18]. Systematic instruction with supplemental task analysis techniques, prompting, and time delay constitutes an effective teaching package based on behaviorism to address students' individualized needs [18]. The literature reports that students develop higher cognitive skills from constructivism approaches and especially inquiry-based learning in science classes; nevertheless, few studies have focused on this approach for students with IDs [18,19].

Furthermore, integrating various digital solutions into the teaching process can enhance students' academic performance. These include video-based instruction [20], Computer Assisted Instruction (CAI), video prompting for teaching scientific terminology [21], and adapted e-texts with multimedia and hyperlinks to enhance comprehension [22]. Additionally, immersive technologies promise engagement, motivation, and positive learning outcomes in teaching scientific concepts to students with ID [23]. Incorporating Digital Learning Objects (DLOs) into teaching can enhance vocabulary acquisition while sustaining high levels of student engagement [24,25].

A Digital Learning Object is defined as a "small, self-contained, reusable, and pedagogically complete structure of learning content. . . aimed at delivering learning experiences" [26] (p. 257). DLOs' key characteristics include reusability, granularity, discoverability, accessibility, interoperability, adaptability, durability, generativity, and manageability, and they are stored in online repositories [27]. Several studies have highlighted the positive impact of DLOs in special education and inclusion [28–30]. Recently, Mallidis-Malessas and colleagues [25] investigated the impact of DLOs from the Hellenic open repository Photodentro on learning grade-aligned Physics concepts, specifically transverse waves and the simple pendulum concept. The study revealed improved learning outcomes and positive student experience.

Given the limited literature on health literacy for vulnerable populations, this study focuses on designing an empirical investigation to support digital health literacy among students with ID, helping them navigate health-related challenges in everyday life. This work contributes to the "key steps to equity and inclusion in education" proposed recently by the Organization for Economic Co-operation and Development (OECD), particularly by supporting teachers in developing essential competencies and knowledge areas, as well as fostering and monitoring students' progress [31].

This study employed a single-subject design [32,33] to investigate the contribution of a digitally supported educational intervention on students with ID through a series of lessons on the cognitive and affective determinants of health during an epidemic or pandemic. Students' achievements included the acquisition of scientific terminology related to infectious diseases and their characteristics. The following research question was posed: What is the effect of the digitally supported educational intervention on students' ability to acquire health literacy content on infectious diseases, focusing on the COVID-19 pandemic?

2. Materials and Methods

2.1. Participants and Setting

Twenty students from a public special vocational high school in a northern city of Greece participated. Eleven male and nine female students aged 18–25 years were chosen from 10 special education classrooms (each classroom accommodated three to eight students). Students' inclusion criteria were: (a) a diagnosis of ID from a government agency, (b) adequate visual discrimination to select words from an array on the computer screen, (c) motor ability to independently use a computer mouse to select stimuli from an array on a computer screen, and (d) verbal response. All 20 participants followed an individualized

education program to increase their vocabulary, reading fluency, and comprehension. The students required prompting and oral guidelines to complete the activities. Regarding the participation protocols, informed consents were obtained from the parents. To safeguard the privacy of all participants, pseudonyms were assigned. An information sheet was provided for participants, which contained information about the research team, the objectives of the study, and the rights and responsibilities assumed through participation. Students were also informed about their voluntary participation, the possibility to withdraw from the study at any time, and the anonymity of the questionnaires used.

All sessions were conducted in the computer lab by one of the authors, the students' Physics teacher, in a one-to-one format. Additionally, two members of the teaching staff helped students collect reliability data.

2.2. Materials

The intervention was based on the educational scenario titled "Cognitive and affective determinants of health during an epidemic/pandemic outbreak for students with Intellectual Disabilities" found at the open access Photodentro PAFSE repository (https://photodentro.pafse.eu/handle/8586/176?&locale=en (accessed on 20 August 2024)). The scenario is supported by four Digital Learning Objects, which can also be found in the open repository Photodentro PAFSE. Additionally, the scenario involves other open digital educational resources, such as videos, which have been retrieved from open online resources (A Day at School, Unicef: https://www.unicef.org/greece/en/stories/day-school-during-covid-19 (accessed on 20 August 2024), or All about Coronavirus: A Video for Kids and Their Families | University of Michigan School of Public Health: https://www.youtube.com/watch?v=6lJQ123_4e8 (accessed on 20 August 2024)). The authors developed both the scenario and the DLOs as parts of the HORIZON 2020 "Partnerships for Science Education—PAFSE" project (https://pafse.eu/ (accessed on 20 August 2024)). Table 1 shows the structure of the scenario.

The DLOs included three interactive concept maps and one interactive infographic. To ensure content validity, the DLOs were designed and developed according to the principles of the Universal Design Framework, confirming accessibility for all students with ID [34,35]. The concept map "COVID-19" aimed to teach students with ID to apply vocabulary regarding coronavirus (Figure 1). This DLO consists of three propositions presented successively, incorporating task analysis steps to align with this instructional technique for students with disabilities. Students select the given words or phrases to complete each proposition. As each proposition is completed, the DLO provides immediate feedback and reinforcement.

The two other concept maps also followed the same design principles (Figure 2). Their content concerns the symptoms (http://photodentro.pafse.eu/handle/8586 /41 (accessed on 20 August 2024)) and the transmission of COVID-19 (http://photodentro.pafse.eu/handle/8586/42 (accessed on 20 August 2024)), respectively.

The fourth DLO is an interactive infographic titled "COVID-19" (Figure 3), which focuses on measures to slow the transmission of infectious diseases, emphasizing the COVID-19 pandemic. It also aims to enhance decision-making skills during a pandemic outbreak. The infographic consists of twelve frames, colored in green (protective measures) or red (measures against protection), aligned with corresponding images. Students have to select the images to complete the frames and receive feedback and reinforcement.

Table 1. The structure of the educational scenario.

Scenario Title Main partner responsible Overview/context Scientific content and its relevance to public health education Estimated duration
STEM content Content glossary Pedagogical glossary
Competences/learning goals Classroom organization requirements Prerequisite knowledge and skills
School research project Teacher guidance notes/ professional development actions Assessment activities
Digital Learning Objects (DLOs) I. Concept map COVID-19 (http://photodentro.pafse.eu/handle/8586/40 (accessed on 10 August 2024)) II. Concept map of symptoms COVID-19 (http://photodentro.pafse.eu/handle/8586/41 (accessed on 10 August 2024)) III. Concept map of the transmission of COVID-19 (http://photodentro.pafse.eu/handle/8586/42 (accessed on 10 August 2024)) IV. Infographic COVID-19 (http://photodentro.pafse.eu/handle/8586/43 (accessed on 4 June 2023)) **Supplementary educational resources (SERs)** https://www.rch.org.au/ccch/covid-19/ (accessed on 10 August 2024) (translated infographic about COVID-19) https://www.youtube.com/watch?v=MVvVTDhGqaA (accessed on 10 August 2024) (translated video about COVID-19, Eurac Research) https://www.unicef.org/greece/en/stories/day-school-during-covid-19 (accessed on 10 August 2024) (video A Day at school, Unicef) https://www.youtube.com/watch?v=6lJQ123_4e8 (accessed on 10 August 2024) (All about Coronavirus: A Video for Kids and Their Families | University of Michigan School of Public Health) https://www.youtube.com/watch?v=GFm45J8d7HI (accessed on 10 August 2024) (video about viruses)
Teacher–learning activities Principal target
Training phase Lesson 1 (Baseline: orientation phase) Lessons 2, 3 (Intervention: main inquiry—conceptualization—investigation) Lesson 4 (Intervention: applying new knowledge and skills) Lessons 5–6 (Open schooling event: applying new knowledge and skills—conclusions)
Supplementary learning activities I. Video watching and discussion II. Handcrafting
Indicative literature

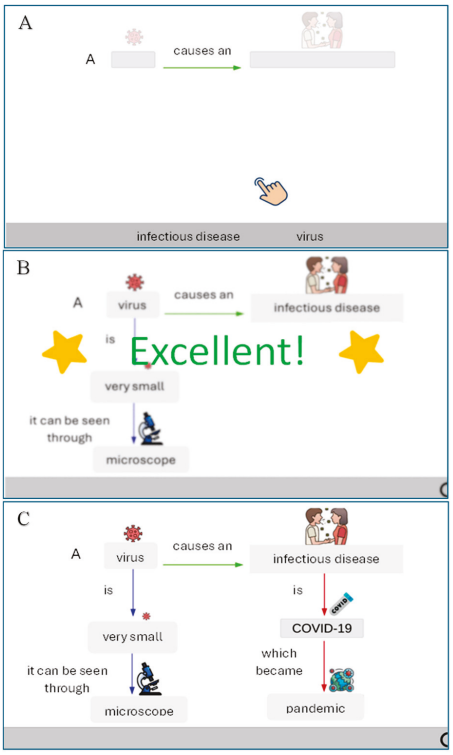


Figure 1. Screenshots (**A**–**C**) of the COVID-19 concept map made with Unity (http://photodentro.pafse.eu/handle/8586/40 (accessed on 20 August 2024)). Reinforcements include yellow star response prompts.

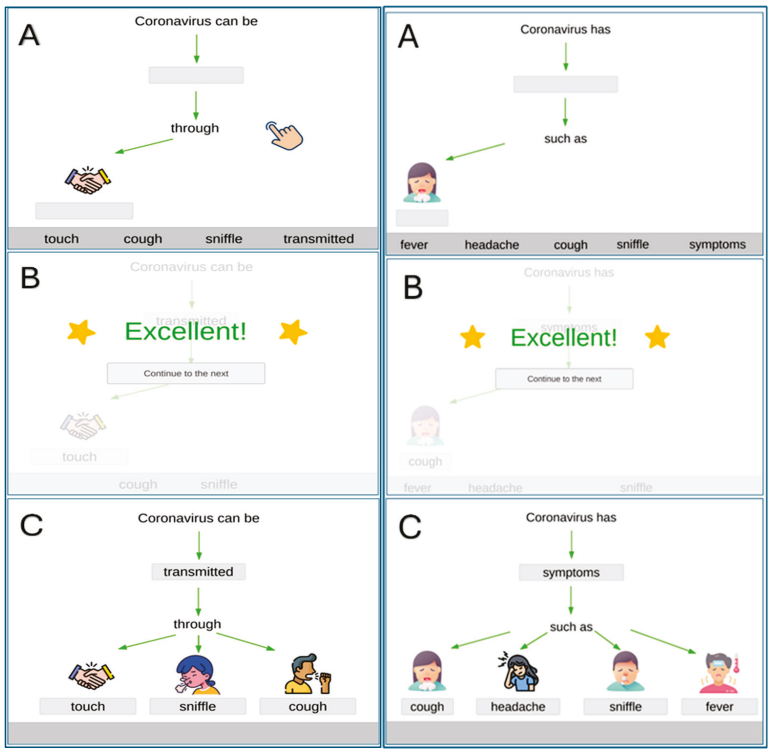


Figure 2. Screenshots (**A**–**C**) of the COVID-19 transmission concept map (**left**) (http://photodentro.pafse.eu/handle/8586/42 (accessed on 20 August 2024)) and COVID-19 symptoms concept map (**right**) (http://photodentro.pafse.eu/handle/8586 /41 (accessed on 20 August 2024)) made with Unity. (**A**–**C**) The COVID-19 concept map was made with Unity (http://photodentro.pafse.eu/handle/8586/40 (accessed on 20 August 2024)). Reinforcements include yellow star response prompts.

Figure 3. Screenshots (**A**–**D**) of the COVID-19 interactive infographic made with Unity (http://photodentro.pafse.eu/handle/8586/43 (accessed on 20 August 2024)). Reinforcements include yellow star response prompts.

2.3. Procedures

2.3.1. Experimental Design

An AB single-subject design was employed to investigate the effects of the inquiry-based intervention on students' performance. "A" refers to the baseline phase, while "B" represents the intervention phase, allowing for the documentation of causal or functional relationships between independent and dependent variables [36]. Experimental control was demonstrated by a clear increase in level and trend from baseline to intervention phases, observed across all students. Figure 4 describes the training, baseline, and intervention phases. Each phase consisted of four sessions to determine the existence of a functional relation between the intervention (independent variable) and student performance (dependent variable). The last session, the open schooling event, acted as a generalization phase.

The AB single-subject design is considered to be suitable for examining individualized intervention effects, particularly in diverse populations, such as students with intellectual disabilities, where variable characteristics in response to interventions are common. This research design enables a detailed assessment of each participant's progress, providing a more precise understanding of how the intervention may impact individual learning outcomes [32,33,36]. Through visual analysis, the risk of overlooking important individual differences can be reduced. Given the heterogeneity within the intellectual disabilities population, a single-subject design, such as an AB design, is particularly valuable for capturing the intervention's effects on each participant. Although the sample of this study is small (20 students), the intra-individual differences of the participants may lead to significant results, even to generalizable findings, and provide replicability.

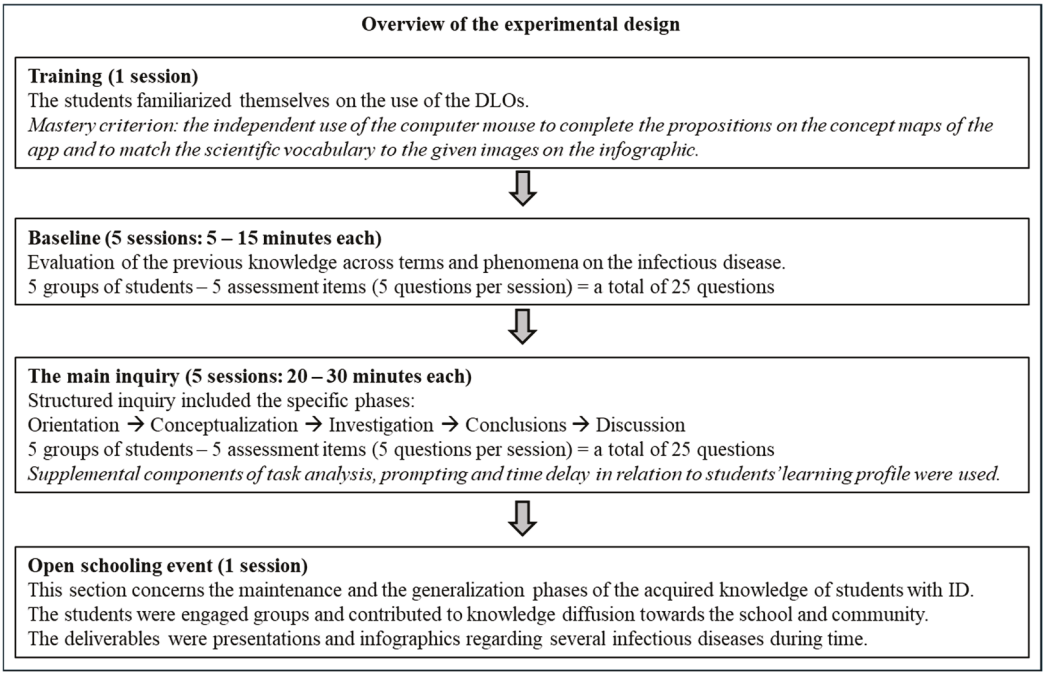


Figure 4. The procedure followed in the empirical study.

2.3.2. Training Phase

A training session was conducted to assess whether students met the mastery criterion for using the DLOs. Simple task analysis steps were implemented to observe students' ability to (a) complete the propositions on the concept maps and (b) match scientific vocabulary to the corresponding images. The researcher confirmed each student's mastery of the DLO and explained how the feedback functions. Within 15 minutes, all students demonstrated proficiency in following the task analysis steps and completing the prepositions on the computer screen.

2.3.3. Baseline Phase

The researcher gathered baseline data over five sessions to assess the prior knowledge of the 20 students. The baseline phase ended when all sessions covering infectious disease descriptions, symptoms, and prevention measures were completed. Oral task directions were provided to each student with identical verbal prompts, such as "choose the correct answer", without providing any additional prompting or feedback.

2.3.4. Digital Technology Supported Inquiry-Based Intervention Phase

Following the baseline phase, the intervention sessions began and lasted for five weeks, with one session per week. Each session lasted from 20 to 30 minutes, depending on each student's level of engagement in the inquiry process. When students entered each session, the researcher launched the Photodentro PAFSE repository, and the students sat in front of their screens. The students interacted with the DLOs and received immediate feedback. Probe sessions with assessment sheets were conducted after the completion of each intervention session. The targeted scientific terminology included terms like virus, infectious disease, incubation time, vaccine, symptoms, and precautionary measures. The main steps of the inquiry process (i.e., orientation, conceptualization, investigation, conclusion, and discussion—applying new knowledge and skills) were designed to help students construct new knowledge as active learners regarding the infectious disease [37–39].

Students received explicit instructions and verbal directions at each step of the structured inquiry to support their learning. The main question, "What is an infectious disease?"

guided the orientation step. Additionally, two brief videos on the COVID-19 pandemic were presented (https://www.youtube.com/watch?v=MVvVTDhGqaA (accessed on 4 June 2022), https://www.unicef.org/greece/en/stories/day-school-during-covid-19 (accessed on 4 June 2022)).

2.4. Data Collection and Dependent Variables

The researchers assessed the effects of the independent variable (i.e., the intervention) on the dependent variables (i.e., the number of correct responses) during the probe sessions. After interacting with the DLOs and following verbal guidelines based on task analysis steps within the structured inquiry, students answered 25 application questions (five per session) related to infectious diseases, their description, and protection measures. Students' responses were collected on data sheets during the probe sessions. Only correct responses were plotted on the graph (Figure 5). Table A1 shows the evaluation sheets used to assess students' acquisition of basic vocabulary and their responses to application questions about infectious diseases. Each student should identify the structure of a virus, recognize key risk factors, define relevant concepts, and understand the importance of vaccination in tracking the progress of a pandemic. Additionally, they had to characterize the relationship between precautionary measures and infection and apply their new knowledge in decision-making scenarios.

Interobserver Agreement

Interobserver agreement (IOA) data were collected from the students across all phases. The second observer was trained to collect IOA data during three sessions in the baseline as well as three sessions in the intervention. Agreement was recorded when the researcher and the second observer provided identical responses regarding the implementation of the intervention. IOA was determined by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100. IOA was 100% for all the students.

2.5. Data Analysis

Six features assess changes in outcomes in an AB design: level ("mean score for the data within a phase"), trend ("slope of the best-fitting straight line for the data within a phase"), variability ("standard deviation of data about the best-fitting straight line"), overlap ("proportion of data from one phase that overlaps with data from the previous phase"), immediacy of effect ("change in level between the last three data points in one phase and the first three data points of the next"), and consistency of data patterns in similar phases ("consistency in the data patterns from phases with the same conditions") [40]. Experimental control is demonstrated when the design documents at least three demonstrations of the experimental effect at three different points in time with a single participant [33].

To express the change in academic performance between the baseline and intervention phases, we calculated the Percentage of Nonoverlapping Data (PND) effect size, which determines the effectiveness of the inquiry-based intervention [41]. PND >70% means effective interventions, 50–70% partial effectiveness, and PND <50% shows an ineffective intervention.

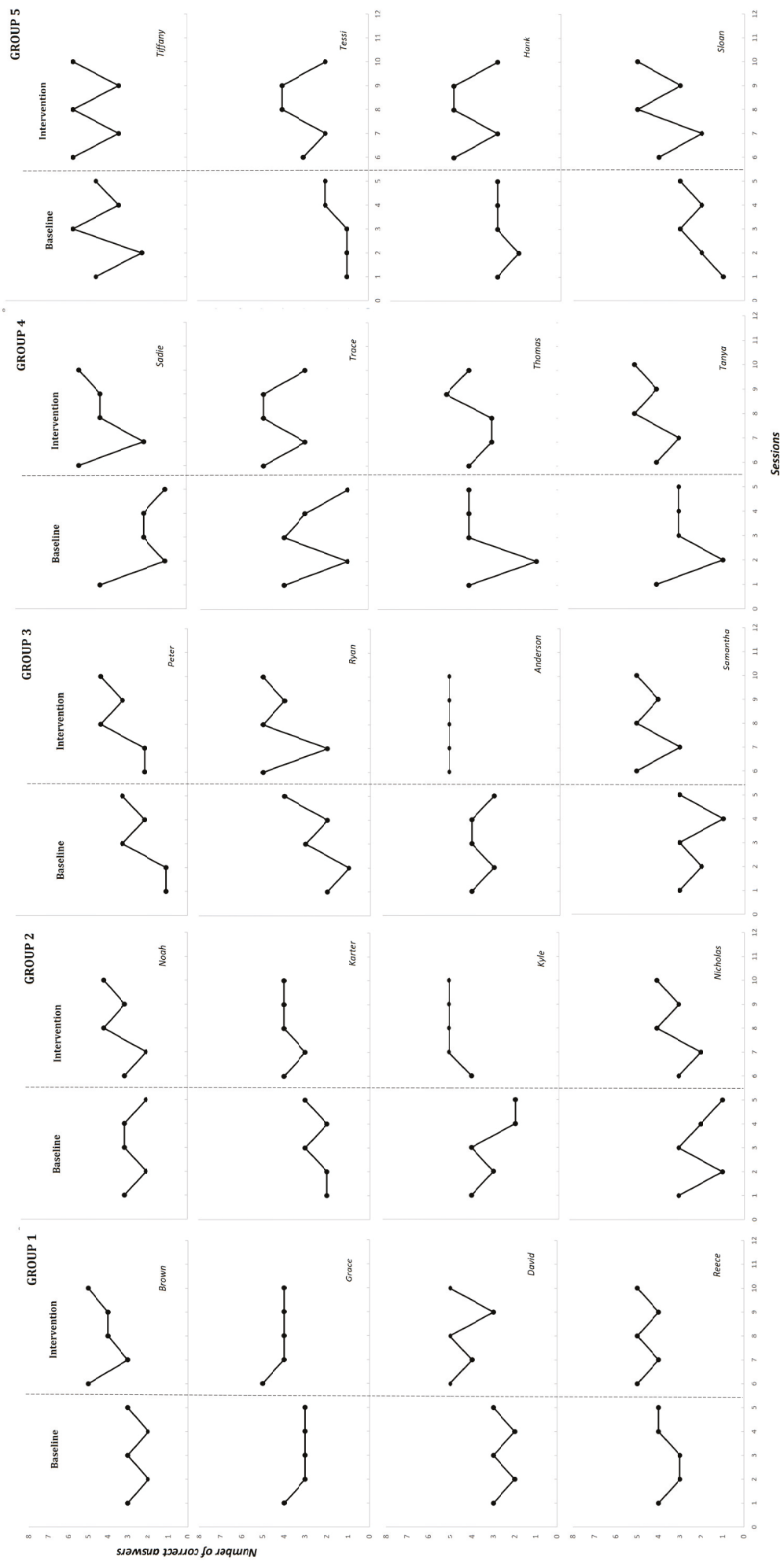


Figure 5. Number of correct responses for the students of the 5 groups participating in the probe sessions. Baseline: 0–5. Intervention: 6–10.

3. Results

Learning Outcomes Across Baseline and Intervention Phases

The visual inspection of the baseline scores indicated that for all students, baseline levels were very low to low and quite variable, with indicative decelerating or stable trends prior to beginning intervention (Table 2 and Figure 5). For example, Brown, David, Noah, Karter, and Anderson had similar data paths during baseline and intervention with the DLOs. For these students, baseline data show initially variable trends in Sessions 1 to 5 (Figure 5), stabilizing or decelerating before intervention. The data of Peter, Ryan, Anderson, Samantha, Sadie, Trace, Thomas, and Tanya showed variability and overlapped the baseline data by 1 to 2 data points (except for Anderson). Grace, Nicholas, Thomas, and Tiffani had almost similar data patterns in the baseline phase and during intervention using DLOs, and the effect sizes for their performance ranged between 0 and 20%. Ryan, Sadie, Trace, and Tanya had distinct data patterns, as their baseline data patterns were variable, ranging from one to four correct answers (Table 2 and Figure 5), and their intervention data patterns changed level; however, they remained variable. Sophie had fewer overlapping data points, while Anderson had no overlapping data points from baseline to intervention conditions.

The 20 students showed increased learning outcomes during the inquiry-based instruction. The functional relationship between dependent and independent variables was robust for most of the students (PND ranged between 60 and 100% for 12 students). The visual analysis of the students' graphs (Figure 5) reveals an upward trend and increased level in the number of correct responses following the introduction of the inquiry-based intervention, compared to the baseline. However, there were data patterns with high variability.

During intervention, 18 students experienced an immediacy of effect and showed an increase in correct responses. All students reached the mastery criterion (completing the basic phases of an inquiry circle). Brown, Grace, David, Reece, Noah, Karter, Kyle, Nicholas, Peter, Ryan, Anderson, and Samantha showed positive learning outcomes at the same rate (Figure 5). When the intervention began, the data were variable yet accelerating or were already stabilized. Brown, David, Karter, Kyle, Anderson, and Samantha had more distinct data patterns. These students had fewer overlapping data points from the baseline phase to the intervention for all sessions than other students. Unlike most students, Anderson began intervention with higher levels (five correct answers in all five sessions). Noah, Nicholas, and Tanya also had distinct data patterns in intervention. In addition, unlike other students, Trace, Hank, and Tessi had distinct data in intervention sessions.

Table 2. Effectiveness data.

Student	Baseline			Intervention			Effect Size
	Mean ± SD (Correct Answers)	Range (Correct Answers)	Visual Analysis (Level, Trend, Variability)	Mean ± SD (Correct Answers)	Range (Correct Answers)	Visual Analysis (Change, Level, Variability)	PND (%)
Brown	2.6 ± 0.5	2–3	low, variable with a slight decelerating trend	4.2 ± 0.8	3–5	abrupt and immediate change in performance, overlapping and variable data	80
Grace	3.2 ± 0.4	3–4	low with a stable trend	4.2 ± 0.4	2–5	immediate change in performance with a stable trend	20
David	2.6 ± 0.5	2–3	low, variable with a stable trend	4.0 ± 1.4	3–5	immediate change in performance overlapping and variable data	80
Reece	3.6 ± 0.5	3–4	low, variable with an overall decelerating trend	4.4 ± 0.9	3–5	variable and overlapping data	60
Noah	2.6 ± 0.5	2–3	low with a stable trend	3.2 ± 0.8	2–4	overlapping data with a slight increasing trend	40
Karter	2.4 ± 0.5	2–3	very low with a stable trend	3.8 ± 0.4	3–4	initial decelerating trend, change in performance, and stable trend	80
Kyle	3.0 ± 1.0	2–4	low, variable with an overall decelerating trend	4.8 ± 0.4	4–5	abrupt and immediate change in level, data with a stable trend	80
Nicholas	2.0 ± 1.0	1–3	very low, variable with an overall decelerating trend	3.2 ± 0.8	2–4	variable data with a slight increasing trend	20
Peter	2.0 ± 1.0	1–3	very low, variable with an increasing trend	3.0 ± 1.0	2–4	variable and overlapping data	40
Ryan	2.4 ± 1.1	1–4	very low, variable with an increasing trend	4.2 ± 1.3	2–5	variable and overlapping data with a slight increasing trend	60
Anderson	3.8 ± 0.8	3–5	low level, with a slight decelerating trend	5.0 ± 0	5–5	very high data with a stable trend	100
Samantha	2.4 ±0.9	1–3	variable	4.4 ± 0.9	3–5	variable and overlapping data with a slight increasing trend	80
Sadie	2.0 ± 1.2	1–5	very low with a decelerating trend	4.0 ± 1.2	2–5	variable and overlapping data	40
Trace	2.6 ± 1.5	1–4	variable with a decelerating trend	4.2 ± 1.1	3–5	variable and overlapping data	60
Thomas	3.4 ± 1.3	1–4	variable with an abrupt change in stable trend	3.8 ± 0.8	3–4	variable and overlapping data with a stable trend	20
Tanya	2.8 ± 1.1	1–4	variable with an accelerating trend	4.2 ± 0.8	3–5	variable and overlapping data with a slight increasing trend	40
Tiffany	3.6 ± 3.6	2–5	variable	4.2 ± 1.1	3–5	variable and overlapping data	0
Tessi	1.4 ± 1.4	1–2	very low with a stable trend	3.0 ± 1.0	2–4	variable and overlapping data	60
Hank	2.8 ± 0.4	2–3	low with a stable trend	4.2 ± 1.1	3–5	immediate change in level, variable data with a stable trend	60
Sloan	2.2 ± 0.8	1–3	very low with an accelerating trend	3.8 ± 1.3	2–5	overlapping data with a slight increasing trend	60

4. Discussion

The purpose of this study was to teach basic content on infectious diseases to students with IDs to acquire health literacy and help them understand the impact of the COVID-19 pandemic on people's lives. Twenty students with mild IDs participated in a structured inquiry intervention supported using DLOs. The results assessing the impact of the intervention on the acquisition of targeted terms and their application showed an increase in the number of correct responses during probes. The twenty students, the teacher, and the two members of the teaching staff agreed that the intervention was effective and expressed a desire to continue using DLOs from the PAFSE Photodentro repository in their teaching process.

The visual analysis of the students' responses revealed that during the baseline phase, all students demonstrated low performance, struggling with the challenging content in Biology and Health Sciences. However, upon entering the intervention, most students exhibited an immediate change in level and demonstrated improved achievements. A functional relationship between the dependent and independent variables was established for 12 students by comparing their performance during the baseline phase with their outcomes after the intervention.

Our results are consistent with previous studies on the contribution of DLOs to teaching academic content in special education [25,28,29,42]. Enhancing the health literacy of students with ID provides a fundamental context for developing meaningful understanding, adopting social behaviors, and communicating ideas related to daily life situations [15,19,21,43].

Regarding students' inquiry skills, an open schooling event with the engagement of students, teachers, parents, and other experts was organized one month after the intervention. The event was used to assess knowledge generalization and communication. Prior to the event, students navigated the provided web resources to identify information on infectious diseases. The researcher encouraged students to pose questions, search in line with their concerns, and explain the process they had to follow (e.g., what will you explore? Define your objective. What do you need to find out? How will you carry out your investigation?). Under these conditions, students were asked to generalize the content they learned by focusing on other infectious diseases. They followed the steps of a simple inquiry process and presented content regarding infectious diseases other than COVID-19. The students communicated the results of their collaborative work. For example, they presented basic information, such as the definition, symptoms, and transmission of Ebola, Yellow Fever, Flu, Cholera, Zika virus, and other pandemics in the past. The majority of the students developed inquiry skills across five dimensions: asking questions, planning and carrying out investigations, analyzing and interpreting data, constructing explanations, and obtaining, evaluating, and communicating information. Specifically, each student (who demonstrated a high PND effect size) could propose certain actions to promote public health, communicate the adoption of choices by others, demonstrate values and adopt individual attitudes that lead to public health, select appropriate information to describe the progress of public health during a pandemic outbreak, identify the community challenges in relation to pandemic outbreak, connect them with social behavior, and find the relevant resources to address them. The other students (Grace, Noah, Nicholas, Peter, Sadie, Thomas, Tanya, and Tiffani presented low scores in PND effect sizes: 0–40) were supported by the teachers and their classmates in order for them to be involved in the inquiry process.

The students agreed that they enjoyed the intervention and reported that the DLOs were easy to use and helped them engage in the learning process. Some indicative comments were: "I liked the lessons about viruses, I will remember this information!", "I liked that we talked about diseases and how we can avoid them", "I liked that we learned about

the coronavirus and the differences between the viruses", "I liked that we could express our opinion about various diseases".

This study demonstrates the possible benefits of our approach for students with IDs. Inquiry-based learning may provide a motivating experience for students with IDs, including student-centered tasks, supported by content adaptations and visual scaffolding that transform the activities accessible to them and improve achievement.

To conclude, this study adds to prior evidence that digital technology makes scientific content accessible for students with IDs and contributes significantly to inquiry skills (science skills or practices, problem solving skills, cognitive and metacognitive skills) that are strengthened during the educational process, especially the inquiry circle [21,43,44]. Considering the specific characteristics of mild ID, the use of DLOs helped students to increase their understanding of infectious diseases. The concept maps consisted of prepositions based on easy-to-read methods, designed to facilitate students in highlighting basic content. The sequential presentation of prepositions on the computer screen, based on task analysis steps, helped students minimize memory overload and process information about the infectious disease [45]. Several visual stimuli (i.e., gestures, such as the pointing finger, Figures 1 and 2) prompted students to choose the given words and complete the propositions. The reinforcements (yellow star response prompts, Figures 1–3) motivated students to sustain their interest and follow the task by reducing the attention deficits or giving up the attempt [16]. Our results are consistent with previous findings showing that students who used DLOs remained engaged and showed strong involvement in the teaching process [46].

Overall, this study agrees with the findings of previous studies [47,48], underlines the significance of health literacy, and draws attention to the importance of increasing the availability of instructional methods such as inquiry-based learning to support students with IDs and improve science and health literacy.

Inclusion is a key implication of this study. Students with IDs can use DLOs developed according to Universal Design principles, follow an educational scenario, and communicate health issues with their fellow students, teachers, and family members, as demonstrated during the open schooling event.

One limitation of this study is that all students entered the intervention phase simultaneously [49], as the completion of an inquiry circle was used as the mastery criterion. A second limitation relates to the varying levels of technical proficiency. The participants were selected based on specific inclusion criteria, such as adequate visual discrimination, motor skills, and verbal responsiveness. A third limitation of this study concerns the variability in baseline data. As observed in Tiffany's results, if baseline data are highly variable, the established threshold (highest or lowest point) may be unrealistic or misleading. This variability could affect the interpretation of intervention effectiveness.

Future studies could explore alternative research designs, expand to include individuals with other disabilities, and investigate different DLOs. For example, future research could complement the findings of this study with a group design to examine broader trends and generalizability. Future research should examine the effectiveness of the DLO intervention across a broader spectrum of technical proficiency, with particular attention to the integration of assistive technology tools to support diverse learner needs. In addition, future research should address the third limitation by collecting additional baseline data to ensure it accurately represents the true pre-intervention state.

Digital health literacy enhances preventive measures and adherence to a healthy lifestyle, boosts skills development, and allows individuals to engage in health-related decisions. We believe that this work contributes to a reduction in educational inequalities as proposed by the OECD's "Equity and Inclusion in Education" [31] and UNESCO's "Ed-

ucation For All (EFA)" (https://unevoc.unesco.org/home/TVETipedia+Glossary/show=term/term=Education+for+all (accessed on 20 August 2024)).

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

Funding: This research was funded by the European Union's Horizon 2020 research and innovation program, grant number No 101006468. The APC was funded by No 101006468.

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

Acknowledgments: The authors would like to thank the director, the teachers, and the students for their participation in the study.

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

Appendix A

Table A1. Questionnaire on infectious diseases.

1. What is a virus? (A) A microscopic plant. (B) A microscopic infectious agent. (C) A small animal.
2. A known virus is coronavirus. Write another virus you know:
3. Why can't we see a virus? (A) Because a virus is very small. (B) Because a virus is always hidden. (C) Because a virus moves very fast.
4. How can we see a virus? We can see a virus through: (A) glasses. (B) a magnifier. (C) a microscope.
5. When someone is under a disease, they appear
6. Recognizing a disease is called
7. Which of the following is an infectious disease? (A) asthma. (B) cancer. (C) COVID-19.
8. How can we call people who get a virus?
9. What are the transmission ways?
10. When a virus causes a disease, this virus is called: (A) pathogenic. (B) infection. (C) host.
11. The incubation time of the virus is the time until (A) the first symptoms appear. (B) go to the doctor. (C) take medicine.
12. The incubation of a disease can take even a few days. True—False
13. A virus cannot be transmitted (A) through air. (B) through physical contact. (C) through a healthy person.
14. When a virus has infected too many people around the world, it is called: (A) an epidemic. (B) a pandemic. (C) nothing like that.
15. Droplets created when someone coughs are a way of transmission. True—False
16. A disease which is spread all over the world is called
17. A lot of viruses come from animals.

Table A1. *Cont.*

18. Fighting a disease is called (A) treatment. (B) prevention. (C) immunity.
19. The defense system of an organism is called: (A) lymphatic. (B) immune. (C) cardiovascular.
20. Which could be symptoms of a virus? (A) sniffle and cough. (B) fever and headache. (C) all of them.
21. How do vaccines work? (A) they strengthen the body's defence/immune system. (B) they cause other diseases. (C) they last forever.
22. How can you protect yourself from being infected by a virus? (A) being vaccinated. (B) visit the doctor. (C) you cannot protect yourself whatever you do.
23. What are the precautionary measures? (A) a doctor diagnosis for a disease. (B) some acts that help prevent diseases. (C) the symptoms of a disease.
24. How can you make decisions to prevent the spread of an infection? (A) social distancing. (B) by the law. (C) there is nothing you can do.
25. An example of social distancing is (A) a pharmacological intervention. (B) a party in a crowded place. (C) hand washing.

References

1. Aishworiya, R.; Kang, Y.Q. Including children with developmental disabilities in the equation during this COVID-19 pandemic. *J. Autism Dev. Disord.* **2021**, *51*, 2155–2158. [CrossRef] [PubMed]
2. Rodgers, R.; Tschida, J. The Impact of COVID-19 on Individuals with Intellectual and Developmental Disabilities: Clinical and Scientific Priorities. *Am. J. Psychiatry* **2020**, *177*, 1091–1093. [CrossRef]
3. Sabatello, M.; Landes, S.D.; McDonald, K.E. People with disabilities in COVID-19: Fixing our priorities. *Am. J. Bioeth.* **2020**, *20*, 187–190. [CrossRef] [PubMed]
4. Courtenay, K.; Perera, B. COVID-19 and people with intellectual disability: Impacts of a pandemic. *Ir. J. Psychol. Med.* **2020**, *37*, 231–236. [CrossRef]
5. Xue, H.; Taylor, L. When do people believe, check, and share health rumors on social media? Effects of evidence type, health literacy, and health knowledge. *J. Health Psychol.* **2023**, *28*, 607–619. [CrossRef]
6. Kuper, H.; Smythe, T. Are People with Disabilities at Higher Risk of COVID-19-Related Mortality? A Systematic Review and Meta-Analysis. *Public Health* **2023**, *222*, 115–124. [CrossRef]
7. Ying, J.; Tan, G.M.Y.; Zhang, M.W. Intellectual Disability and COVID-19: A Bibliometric Review. *Front. Psychiatry* **2023**, *13*, 1052929. [CrossRef]
8. Parker, M.; Alfaro, P. *Education During the COVID-19 Pandemic: Access, Inclusion and Psychosocial Support*; Studies and Perspectives Series-ECLAC Subregional Headquarters for the Caribbean, No. 104; Economic Commission for Latin America and the Caribbean (ECLAC): Santiago, Chile, 2022.
9. Boyle, C.A.; Fox, M.H.; Havercamp, S.M.; Zubler, J. The public health response to the COVID-19 pandemic for people with disabilities. *Disabil. Health J.* **2020**, *13*, 100943. [CrossRef]
10. Delgado, P.; Ávila, V.; Fajardo, I.; Salmerón, L. Training young adults with intellectual disability to read critically on the internet. *J. Appl. Res. Intellect. Disabil.* **2019**, *32*, 666–677. [CrossRef]
11. Geukes, C.; Bröder, J.; Latteck, Ä.-D. Health literacy and people with intellectual disabilities: What we know, what we do not know, and what we need: A theoretical discourse. *Int. J. Environ. Res. Public Health* **2019**, *16*, 463. [CrossRef]
12. Chadwick, D.D.; Ågren, K.A.; Caton, S.; Chiner, E.; Danker, J.; Gómez-Puerta, M.; Heitplatz, V.N.; Johansson, S.; Normand, C.L.; Murphy, E.; et al. Digital inclusion and participation of people with intellectual disabilities during COVID-19: A rapid review and international bricolage. *J. Policy Pract. Intellect. Disabil.* **2022**, *19*, 307–324. [CrossRef]
13. Choukou, M.-A.; Sanchez-Ramirez, D.C.; Pol, M.; Uddin, M.; Monnin, C.; Syed-Abdul, S. COVID-19 infodemic and digital health literacy in vulnerable populations: A scoping review. *Digit. Health* **2022**, *8*, 20552076221076927. [CrossRef] [PubMed]
14. National Research Council. *Next Generation Science Standards: For States, by States*; The National Academies Press: Washington, DC, USA, 2013.

15. Roberts, C.A.; Kim, S.; Tandy, J.; Meyer, N. Using content area literacy strategies during shared reading to increase comprehension of high school students with moderate intellectual disability on adapted science text. *Educ. Train. Autism Dev. Disabil.* **2019**, *54*, 147–160.
16. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; American Psychiatric Association: Arlington, VA, USA, 2013.
17. Individuals with Disabilities Education Improvement. 20 U.S.C. §1400, H.R. 1350. Congressional Record, 17 November 2004, pp. H9895–9959. Available online: https://www.congress.gov/bill/108th-congress/house-bill/1350/all-actions (accessed on 28 April 2025).
18. Knight, V.F.; Wood, L.; McKissick, B.R.; Kuntz, E.M. Teaching science content and practices to students with intellectual disability and autism. *Remedial Spec. Educ.* **2020**, *41*, 327–340. [CrossRef]
19. Rizzo, K.; Taylor, J. Effects of inquiry-based instruction on science achievement of students with disabilities: An analysis of the literature. *J. Sci. Educ.* **2016**, *19*, 2. [CrossRef]
20. Park, J.; Bouck, E.; Duenas, A. The effect of video modeling and video prompting interventions on individuals with intellectual disability: A systematic literature review. *J. Spec. Educ. Technol.* **2019**, *34*, 3–16. [CrossRef]
21. McKissick, B.; Davis, L.; Spooner, F.; Fischer, L.; Graves, C. Using computer-assisted instruction to teach science vocabulary to students with autism spectrum disorder and intellectual disability. *Rural Spec. Educ. Q.* **2018**, *37*, 207–218. [CrossRef]
22. Wood, L.; Browder, D.M.; Spooner, F. Teaching listening comprehension of science e-texts for students with moderate intellectual disability. *J. Spec. Educ. Technol.* **2020**, *35*, 272–285. [CrossRef]
23. Mikropoulos, T.; Iatraki, G. Digital technology supports science education for students with disabilities: A systematic review. *Educ. Inform. Technol.* **2022**, *28*, 3911–3935. [CrossRef]
24. Iatraki, G.; Mallidis-Malessas, P.; Mikropoulos, T.A. Digital learning objects support grade-aligned Physics instruction for high school students with mild intellectual disability. In Proceedings of the 9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion, Lisbon, Portugal, 31 August–2 September 2022; ACM: Lisbon, Portugal, 2020; pp. 213–218.
25. Mallidis-Malessas, P.; Iatraki, G.; Mikropoulos, T.A. Teaching physics to students with intellectual disabilities using digital learning objects. *J. Spec. Educ. Technol.* **2021**, *37*, 61–74. [CrossRef]
26. Topali, P.; Mikropoulos, T. Digital Learning Objects for Teaching Computer Programming in Primary Students. In *Technology and Innovation in Learning, Teaching and Education*; Tsitouridou, M., Diniz, A.J., Mikropoulos, T., Eds.; Springer: Basel, Switzerland, 2019; pp. 256–266.
27. Megalou, E.; Kaklamanis, C. Photodentro LOR, the Greek national learning object repository. In *INTED Proceedings, 8th International Technology, Education and Development Conference, 10–12 March 2014, Valencia, Spain*; Gómez Chova, L., López Martínez, A., Candel Torres, I., Eds.; IATED Academy: Valencia, Spain, 2014; pp. 309–319.
28. Benigno, V.; Capuano, N.; Mangione, G.R. A Web-Based Knowledge Hub for Special and Inclusive Education. *Int. J. Emerg. Technol. Learn.* **2015**, *10*, 5–13. [CrossRef]
29. Jadán-Guerrero, J.; Guerrero, L.A. A virtual repository of learning objects to support literacy of SEN children. *IEEE Rev. Iberoam. Tecnol. Aprendiz.* **2015**, *10*, 168–174. [CrossRef]
30. Roman, J.; Mehta, D.; Saija, P.S. Learning needs for special needs learners: A graph-based adaptive approach for content sequencing. In Proceedings of the International Conference on Soft Computing and its Engineering Applications, Anand, India, 1–2 December 2017; IEEE: London, UK, 2017; pp. 1–5.
31. OECD. *Equity and Inclusion in Education: Finding Strength Through Diversity*; OECD Publishing: Paris, France, 2023. [CrossRef]
32. Gast, L.D.; Ledford, R.J. *Single Case Design Methodology: Applications in Special Education and Behavioral Sciences*; Routledge: New York, NY, USA, 2014.
33. Horner, R.H.; Carr, E.G.; Halle, J.; McGee, G.; Odom, S.; Wolery, M. The use of single-subject research to identify evidence-based practice in special education. *Exceptionality* **2005**, *71*, 165–179. [CrossRef]
34. AlRawi, J.M.; AlKahtani, M.A. Universal design for learning for educating students with intellectual disabilities: A systematic review. *Int. J. Dev. Disabil.* **2021**, *47*, 141–154. [CrossRef]
35. Coyne, P.; Pisha, B.; Dalton, B.; Zeph, L.; Smith, S. Literacy by design: A universal design for learning approach for students with significant intellectual disabilities. *Remedial Spec. Educ.* **2012**, *33*, 162–172. [CrossRef]
36. Hammond, D.; Gast, D. Descriptive analysis of single subject research designs: 1983–2007. *Educ. Train. Autism Dev. Disabil.* **2010**, *45*, 187–202.
37. Constantinou, C.P.; Tsivitanidou, O.E.; Rybska, E. What Is Inquiry-Based Science Teaching and Learning? In *Professional Development for Inquiry-Based Science Teaching and Learning*; Tsivitanidou, O., Gray, P., Rybska, E., Louca, L., Constantinou, C., Eds.; Springer: Cham, Switzerland, 2018; pp. 1–19. [CrossRef]
38. McDermott, L.C.; Shaffer, P.S.; Constantinou, C.P. Preparing teachers to teach physics and physical science by inquiry. *Phys. Educ.* **2000**, *35*, 411. [CrossRef]

39. Pedaste, M.; Mäeots, M.; Siiman, L.; De Jong, T.; Van Riesen, S.; Kamp, E.; Manoli, C.; Zacharia, Z.; Tsourlidaki, E. Phases of Inquiry-Based Learning: Definitions and the Inquiry Cycle. *Educ. Res. Rev.* **2015**, *14*, 47–61. [CrossRef]
40. Kratochwill, T.R.; Hitchcock, J.; Horner, R.H.; Levin, J.R.; Odom, S.L.; Rindskopf, D.M.; Shadish, W.R. Single-Case Designs Technical Documentation. What Works Clearinghouse. 2010. Available online: https://files.eric.ed.gov/fulltext/ED510743.pdf (accessed on 1 February 2025).
41. Scruggs, T.E.; Mastropieri, M.A. How to summarize single-participant research: Ideas and applications. *Exceptionality* **2001**, *9*, 227–244. [CrossRef]
42. de Macedo, C.M.S.; Ulbricht, V.R. Universal design and accessibility standards in online learning objects. In *Universal Access in Human-Computer Interaction*; Stephanidis, C., Antona, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8011, pp. 157–166. [CrossRef]
43. McMahon, D.; Cihak, D.; Wright, R.; Bell, S. Augmented reality for teaching science vocabulary to postsecondary education students with intellectual disabilities and autism. *J. Res. Technol. Educ.* **2016**, *48*, 38–56. [CrossRef]
44. Ciullo, S.; Falcomata, T.; Pfannenstiel, K.; Billingsley, G. Improving learning with science and social studies text using computer-based concept maps for students with disabilities. *Behav. Modif.* **2015**, *39*, 117–135. [CrossRef] [PubMed]
45. Nurmi, S.; Jaakkola, T. Promises and Pitfalls of Learning Objects. *Learn. Media Technol.* **2006**, *31*, 269–285. [CrossRef]
46. Natsis, A.; Hormova, H.; Mikropoulos, T.A. Students' views on different learning objects types. In *INTED Proceedings, 8th International Technology, Education and Development Conference, 10–12 March 2014, Valencia, Spain*; Gómez Chova, L., López Martínez, A., Candel Torres, I., Eds.; IATED Academy: Valencia, Spain, 2014; pp. 2363–2372.
47. Turnbull, H.; Dark, L.; Carnemolla, P.; Skinner, I.; Hemsley, B. A Systematic Review of the Health Literacy of Adults with Lifelong Communication Disability: Looking Beyond Accessing and Understanding Information. *Patient Educ. Couns.* **2023**, *106*, 151–162. [CrossRef] [PubMed]
48. Yeap, Z.; Keeley, J.; Skoss, R.; Hunt, S.; Kickett, A.; Saldaris, J.; Nevill, T.; Downs, J. Online Health Literacy Resources for People with Intellectual Disability: Protocol for a Grey Literature Scoping Review. *BMJ Open* **2024**, *14*, e088509. [CrossRef]
49. Tincani, M.; Travers, J. Publishing single-case research design studies that do not demonstrate experimental control. *Remedial Spec. Educ.* **2018**, *39*, 118–128. [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.

computers

MDPI

Article

Students Collaboratively Prompting ChatGPT

Maria Perifanou * and Anastasios A. Economides

SMILE Lab, University of Macedonia, 54636 Thessaloniki, Greece; economid@uom.gr
* Correspondence: mariaperif@gmail.com

Abstract: This study investigated how undergraduate students collaborated when working with ChatGPT and what teamwork approaches they used, focusing on students' preferences, conflict resolution, reliance on AI-generated content, and perceived learning outcomes. In a course on the Applications of Information Systems, 153 undergraduate students were organized into teams of 3. Team members worked together to create a report and a presentation on a specific data mining technique, exploiting ChatGPT, internet resources, and class materials. The findings revealed no strong preference for a single collaborative mode, though *Modes #2, #4*, and *#5* were marginally favored due to clearer structures, role clarity, or increased individual autonomy. Students reasonably encountered initial disagreements (averaging 30.44%), which were eventually resolved—indicating constructive debates that improve critical thinking. Data also showed that students moderately modified ChatGPT's responses (50% on average) and based nearly half (44%) of their overall output on AI-generated content, suggesting a balanced yet varied level of reliance on AI. Notably, a statistically significant relationship emerged between students' perceived learning and actual performance, implying that self-assessment can complement objective academic measures. Students also employed a diverse mix of communication tools, from synchronous (phone calls) to asynchronous (Instagram) and collaborative platforms (Google Drive), valuing their ease of use but facing scheduling, technical, and engagement issues. Overall, these results reveal the need for flexible collaborative patterns, more supportive AI use policies, and versatile communication methods so that educators can apply collaborative learning effectively and maintain academic integrity.

Keywords: ChatGPT; collaborative learning; collaborative modes; collaborative prompting; GenAI; project-based learning; small group; team-based learning

1. Introduction

Recently, generative artificial intelligence (GenAI) tools have gained increasing popularity. They are based on large language models (LLMs) trained on vast amounts of data and can generate human-like content (e.g., text, images, video) in response to prompts (e.g., [1]). They perform various language-related tasks, such as text generation [1,2], question answering [3], and translation, while comprehending the context to produce responses that resemble human language [4]. Currently, the most popular GenAI tool is ChatGPT [5]. It is an LLM based on generative pre-trained transformer (GPT) architecture which is designed for text generation and language understanding (e.g., [6]). In February 2025, there were 3.9 B visits to chatgpt.com [7].

Education has adopted such GenAI tools because they have the potential to facilitate adaptive and personalized learning [2,3,6,8]; increase student learning [1] and engagement [2,3]; enable the understanding of complex concepts [8,9]; provide personalized feedback [2,6,8–10]; aid in developing innovative assessments [6]; and support research

and data analysis [2,4,6]. However, there are concerns with respect to academic integrity, plagiarism, and cheating [1,3,10,11]; potential biases and falsified information [1,2,6]; copyright [2]; privacy [9,11,12]; and the possible reduction in human interaction [13].

An effective instructional method is collaborative learning [14]. Collaborative learning refers to learners collaborating with other learners to co-construct knowledge through social interaction and is based on the theoretical framework of social constructivism learning theory and group or social cognition [14,15]. It involves two or more students working together to achieve a common goal or reach a common objective, typically through peer-directed interactions [16,17]. It is well known [18] that students who learn in small groups typically exhibit higher academic achievement, more positive attitudes toward learning, and more persistence in science, mathematics, engineering, and technology (SMET) courses or programs than their peers who learn individually (not collaboratively). Also, collaboration has significant positive effects on students' knowledge gain, skill acquisition, and perceptions in computer-based learning settings, while computer use has positive effects on knowledge gain, skill acquisition, student perceptions, group task performance, and social interaction in collaborative learning settings [14]. Although various digital technologies have been used in collaborative learning, previous studies pointed out that there is a need for empirical research regarding the use of AI and ChatGPT in collaborative learning [12,19,20].

2. Theoretical Background

Many previous studies argued that artificial intelligence (AI) can facilitate collaborative learning [10,12,21,22]. More specifically, AI can support group creation [12,23], group discussions [2,8,22], collaborative writing [2,24], group outcomes' evaluation [12], and group management [22].

Recent literature reviews found that AI tools support collaborative learning by providing peer learning opportunities [25], enabling personalized learning through feedback and group work [26], and enhancing students' cognitive engagement [27]. Most previous studies used AI-driven learning analytics to analyze students' communicative discourse, behavioral, and evaluation data to display visualizations of students' outcomes [27]. Also, ref. [28] found that using AI to analyze discussion data helps to understand the online collaborative learning process leading to increased students' learning and higher-order presence. Additionally, ref. [29] observed that ChatGPT-supported collaborative argumentation significantly enhanced English as a Foreign Language (EFL) students' argumentative speaking performance, critical thinking awareness, and collaboration tendency. Furthermore, [30] found that AI-empowered recommendations could significantly improve collaborative knowledge-building, co-regulated behaviors, and group performance. On the other hand, a few studies noticed that the overuse of ChatGPT may decrease students' creative and collaborative abilities [31]. For example, when working on group projects, students preferred using ChatGPT on their own rather than brainstorming and working with peers.

Many researchers [12,19,20] recognized the lack of empirical studies on AI in collaborative learning and stressed the need for further research on how AI can be used in collaborative learning. Although several previous studies used AI to support the collaborative learning process, there is limited research on how students actually collaborate and communicate among themselves when they use AI tools. Knowing how students prefer to work together, when prompting ChatGPT, is important because their comfort with a specific collaborative approach may influence their motivation, engagement, and overall learning outcomes. Also, understanding how teammates like to collaborate, when working with ChatGPT, enables instructors to structure collaborative activities that align with their preferences, maximizing participation and knowledge sharing. In a previous

study [32], thirty graduate students reported their preferences regarding different ways that they communicated and collaborated when working with GenAI tools. The current study extends that research by examining a larger sample of undergraduate students who collaboratively designed prompts for ChatGPT. It also analyzes the students' experiences using both quantitative analysis and qualitative analysis (content analysis reporting the frequencies of themes). Furthermore, it examines the initial disagreements in teams, the teams' reliance on AI-generated content, and the students' perceived learning. Thus, this study aims to investigate the following research questions:

RQ1 : *Which collaborative mode do team members prefer when they interact with ChatGPT? What are the pros and cons of each collaborative mode?*

RQ2 : *To what extent do students modify ChatGPT's responses? To what extent do students use ChatGPT's responses in their final output?*

RQ3 : *To what extent do students learn during their collaborative activities?*

RQ4 : *(i) Which communication channel do team members use to collaborate? (ii) What aspects of the communication methods do they like? (iii) What challenges do they face in communication? (iv) What suggestions do they have for improving communication?*

3. Methodology

3.1. Triangulation and Mixed-Methods Research Design

Triangulation was employed to enhance the analysis and interpretation of the collected data by utilizing multiple data sources, researchers, and research methods [33,34]. This approach enhances trust in the data, reduces analyst bias, and provides diverse perspectives, leading to a more comprehensive understanding of the investigated issue. It also reveals insights that might remain hidden when relying on a single method. This study implements between-method triangulation [35], integrating data triangulation (multiple sources) and methodological triangulation (varied methods).

Additionally, this study adopts a mixed-methods research approach [36]. Student responses were analyzed both quantitatively and qualitatively. Descriptive and inferential statistics were used for the quantitative analysis of students' responses to closed-ended questions, while content analysis [37] was applied for the qualitative analysis of students' responses to open-ended questions, as well as data from observations of teams' presentations, focus group discussions, and teams' reports.

3.2. Data Collection Methods

The data collection methods included students' participation in online questionnaires and focus groups discussions, as well as educators' direct observations of teams' presentations and grading of teams' reports. Thus, the data included students' (i) answers to the closed- and open-ended questions of a questionnaire, (ii) perspectives during focus group discussions, (iii) presentations, (iv) reports, and (v) grades.

3.3. Questionnaire

To gather data, a questionnaire was developed using Google Forms. The questionnaire included 19 questions (items) on students' demographics, evaluation of collaborative modes (described in Section 3.8), extent of utilizing ChatGPT-generated content, evaluation of students' learning, and open-ended questions. The questionnaire was initially reviewed by three researchers of technology-enhanced learning and three students from the University of Macedonia, Thessaloniki, Greece. The questionnaire items were reviewed based on whether they cover the issue to be measured, they made sense, and whether the language was correct, clear, and appropriate. Also, the questionnaire was pilot-tested for the clarity

of instructions, length, and usability. Reviewers' suggestions were incorporated into the questionnaire. So, content validity and face validity were assured. Participants were fully informed about the purpose of this study and voluntarily gave their consent and answered the questionnaire. This study was conducted in accordance with ethical guidelines, having received approval from the University of Macedonia Ethical Committee.

3.4. Focus Group Discussions

During week #j (j = 1, 5), there was a 20-min focus group discussion with students that followed Collaborative Mode #j (j = 1, 5), described in Section 3.8, when interacting with ChatGPT. At the end of each class in week #j (j = 1, 5), the educators triggered discussions with students about their perspectives, the advantages and disadvantages of using Mode #j (j = 1, 5), and their roles and the role of ChatGPT in their interactions.

3.5. Content Analysis

Data from students' responses to open-ended questions, observations of teams' presentations, focus group discussions, and teams' reports were analyzed using inductive content analysis [37]. Inductive content analysis is a qualitative research method aimed at the subjective interpretation of text data through systematic coding and theme identification. Two researchers repeatedly read the data to achieve immersion, followed by word-by-word analysis to derive initial codes from significant phrases [38]. The researchers made notes on their impressions, leading to the emergence of a coding scheme that reflected key concepts. Then, they grouped the codes into categories based on their relationships, forming meaningful clusters. Finally, they calculated the code frequency counts for each identified category. The inter-rater reliability was over 90%. In the case of disagreement, they discussed the issues and reached a consensus.

3.6. Context and Participants

A total of 256 first-year undergraduate students took an elective course on the Applications of Information Systems in the department of Economics at the University of Macedonia, Thessaloniki, Greece, during the spring semester of 2024. This was a first-year undergraduate course on businesses' digital infrastructures, data management and business intelligence, telecommunications and networking, and information systems security and applications in businesses. The requirements of the course included the development of a team (small-group) project that accounts for 25% of the student's final grade. One of the course topics in business intelligence included an overview of data mining techniques (e.g., classification, clustering, associations, sequences, anomaly detection, regression, prediction, forecasting). So, a short overview and the definitions of various data mining techniques were presented during one lecture in March 2024. Then, each team selected a specific data mining technique (among six data mining techniques). The goal of each team was to study and collaboratively develop a report on the specific data mining technique chosen during March and April 2024 and present their teamwork in class during May and June 2024.

Initially, students had to freely form teams of three peers and choose to collaboratively investigate a specific data mining technique. In other words, students self-selected their teammates without any instructor intervention in team formation. The teams' final product was a report and a 20–30-slide presentation on their specific data mining technique, presenting it in class. Each team should study and present an overview of their selected data mining technique. The overview should include the definition, input and output, popular algorithms and available software, potential challenges and limitations, common applications in business, examples and real cases, emerging trends, and references. The students could find relevant information in the course's resources (on the university's learning management system), the educator's presentations, books in the university's library, and the

internet. Students were also required to use the free version of ChatGPT (version GPT-3.5 in spring 2024) and other generative AI tools. They were advised to follow a simple and compact form (adapted to the students' level and work) of the Collaborative Preparation, Implementation, Revision (CoPIR) Methodology [32] (presented in Section 3.7). Then, the teams were randomly assigned to five (5) collaborative modes. These collaborative modes are presented in Section 3.8. The members of a particular team were required to follow the specific collaborative mode assigned to them. The team had to include in their presentation only the final results that they obtained after their conversation with ChatGPT and the discussions among the team members. However, in the appendix of their report, the team had to report all the prompts that they gave to ChatGPT, the answers received, the revised prompts, and so on.

Each team had to assign a Creative Commons license to their report and presentation. The references in the team's presentation must follow APA style. All materials (e.g., text, graphs, tables, images, video, responses of ChatGPT) need to be cited wherever they appear in the presentation and also be listed in the references at the end of the presentation and the report.

During a five (5)-week period in May and June 2024, teams (with a random sequence) presented their work in class. In week #j (j = 1 to 5), there were presentations in class by teams that followed Collaborative Mode #j. At the end of each class, there was also a focus group discussion with all students who followed the particular Collaborative Mode #j. Finally, students were expected to answer an online questionnaire regarding their experience using ChatGPT and collaborating with their team members. The online questionnaire was active on Google Forms for two weeks during June 2024.

3.7. Collaborative Preparation, Implementation, Revision (CoPIR) Methodology

All teams were suggested to follow a simplified condensed version of the Collaborative Preparation, Implementation, Revision (CoPIR) methodology [32] adjusted to the team's work. This CoPIR methodology has the following three phases (Figure 1, Table 1):

(1) *Preparation*: Focus on understanding the work (e.g., assignment, project), creating a plan, and assigning roles.
(2) *Implementation*: Generate and validate content individually, then integrate and improve it as a team.
(3) *Revision*: Review, refine, and finalize the output, ensuring quality and academic integrity.

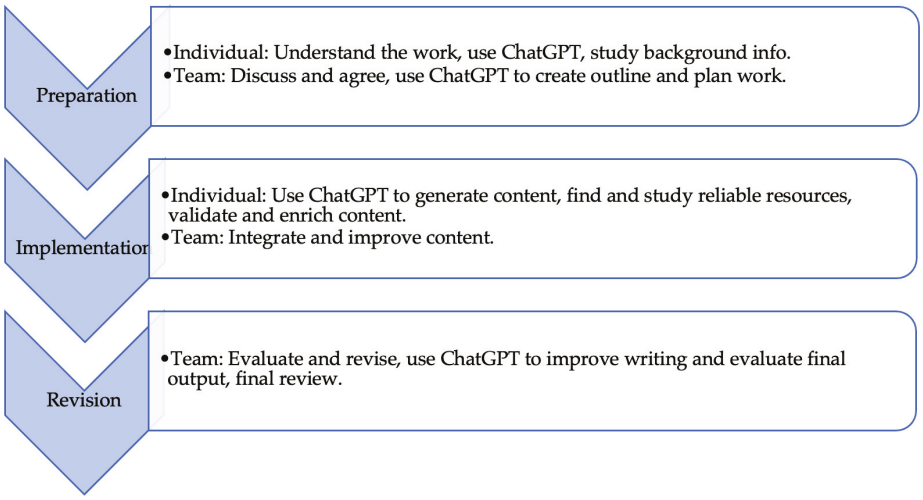


Figure 1. The flow of the Collaborative Preparation, Implementation, Revision (CoPIR) methodology.

Table 1. Collaborative Preparation, Implementation, Revision (CoPIR) methodology for accomplishing work (e.g., assignment, activity, project) with the help of ChatGPT.

Phases	Collaborative Preparation, Implementation, Revision (CoPIR) Methodology	
	Individual Actions	**Team Actions**
1. PREPARATION		
	- Read and understand the requirements, objectives, and deadlines of the work. - Use ChatGPT to explain key concepts, theories, and methods related to the work. - Study and share background information from reliable sources.	- Discuss and agree on the scope, objectives, and expected outcomes of the work. - Use ChatGPT to suggest an outline/structure for the final output and agree on it. - Use ChatGPT to suggest a work plan (roadmap, steps, communication channels, timeline) and assign roles.
2. IMPLEMENTATION		
	- Use ChatGPT to generate content for your part of the work. - Find and study additional reliable resources for your part. - Validate ChatGPT-generated content with reliable sources. - Edit and enrich ChatGPT-generated content with your own findings.	- Integrate, edit, and improve all individual parts of the content collaboratively.
3. REVISION		
		- Evaluate, validate, and revise the integrated content for accuracy and clarity. - Use ChatGPT to improve writing (grammar, syntax, clarity, coherence) and evaluate the final output. - Review and revise the final output, ensuring academic integrity (citations). - Use ChatGPT to create and refine the presentation.

Note that work can correspond to any of the following: assignment, report, presentation, simulation, experiment, portfolio, project, etc.

3.8. Collaborative Prompting When Using GenAI Tools

Team members were required to collaborate with one another following one specific collaborative mode among the five collaborative modes [32]. The teams were randomly divided into five clusters, and a specific collaborative mode was assigned to each cluster. So, teams in a specific cluster were instructed to employ the assigned collaborative mode.

(1) *Collaborative Mode #1 (one in a row)*: Each team member takes turns refining a prompt for ChatGPT (Figure 2). The first member provides an initial prompt, and all members see ChatGPT's response. The second member revises the prompt based on the response and submits a new one. After seeing the new response, the third member submits their revised prompt. Finally, the team discusses ChatGPT's latest response and agrees on the final result. In summary, each member revises the prompt based on ChatGPT's response to another member's prompt. At the end of an iteration cycle, the team discusses the final response to reach a conclusion.

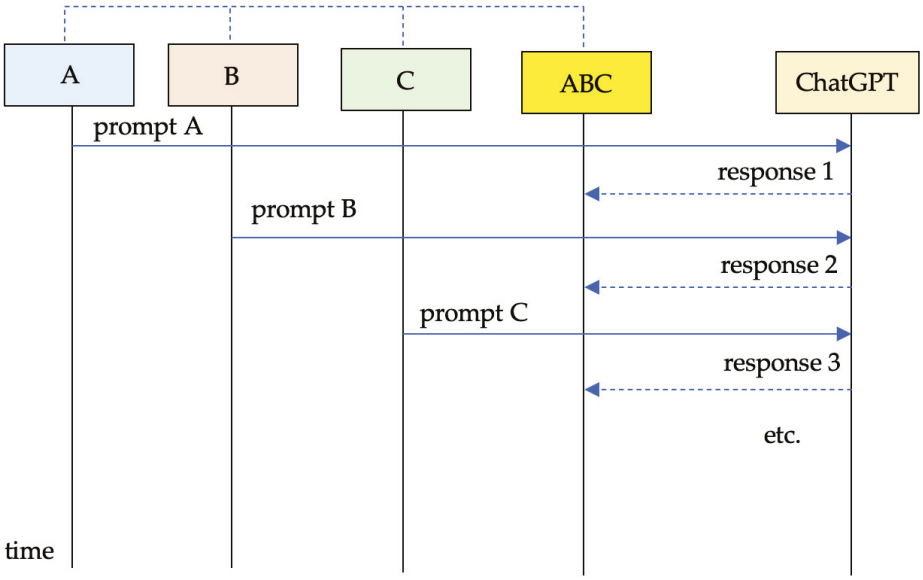


Figure 2. Collaborative *Mode #1* (one in a row) sequence diagram of team ABC (members: A, B, and C) when interacting with ChatGPT.

(2) *Collaborative Mode #2 (team-oriented)*: The team collaborates to create and refine prompts for ChatGPT (Figure 3). They start by agreeing on an initial prompt, which the team leader submits. After seeing ChatGPT's response, they discuss and agree on a revised prompt, which the leader submits again. This process repeats until the team agrees on a final result. This is a simple and collaborative approach where the team works together throughout. In summary, the team always works together to create and refine prompts until they reach a final result.

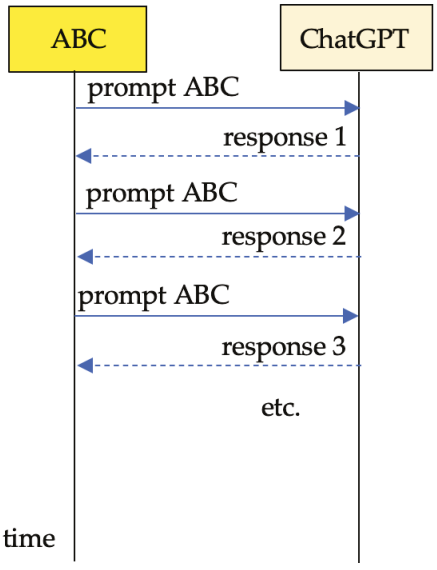


Figure 3. Collaborative *Mode #2* (team-oriented) sequence diagram of team ABC (members: A, B, and C) when interacting with ChatGPT.

(3) *Collaborative Mode #3 (member then team)*: Each team member starts by independently submitting their own prompt to ChatGPT (Figure 4). After reviewing all responses, the team discusses and agrees on a joint revised prompt, which the leader submits. This process is repeated: the team discusses ChatGPT's response, agrees on a new revised prompt, and the leader submits it. Finally, the team reviews the last response and decides on the final result. This approach allows members to explore different ideas independently

before collaborating to refine the prompts and reach a conclusion. In summary, there is an independent start where each member submits their own prompt to explore different perspectives. Then, there is a collaborative iterative refinement where the team discusses responses and agrees on revised prompts until they reach a final result.

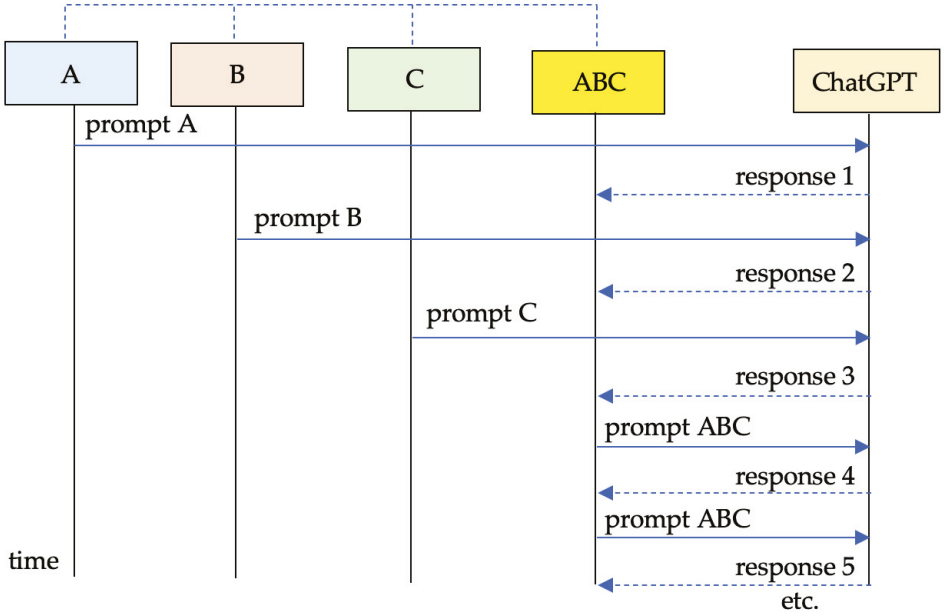


Figure 4. Collaborative *Mode #3* (member then team) sequence diagram of team ABC (members: A, B, and C) when interacting with ChatGPT.

(4) *Collaborative Mode #4 (team to member)*: Each team member starts by independently submitting their own prompt to ChatGPT (Figure 5). After reviewing all responses, the team discusses and suggests revised prompts to each member. Each member then independently decides on their revised prompt and submits it. This process is repeated: the team discusses the new responses and suggests further revisions, and each member independently submits their updated prompt. Finally, the team reviews all responses and agrees on the final result.

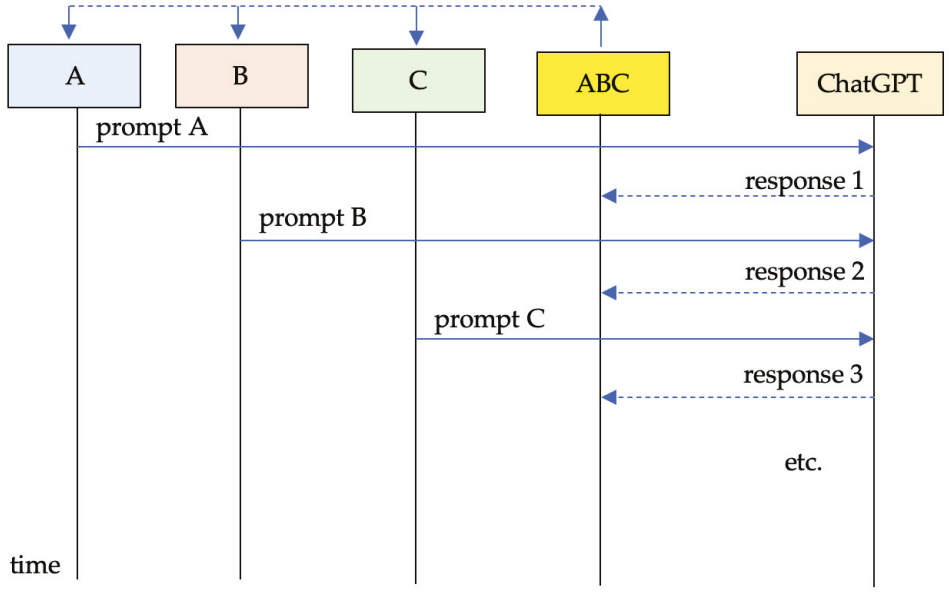


Figure 5. Collaborative *Mode #4* (team to member) sequence diagram of team ABC (members: A, B, and C) when interacting with ChatGPT.

(5) *Collaborative Mode #5 (independently, cooperatively)*: The work is split into three parts. Each team member works independently on their assigned part, using ChatGPT alone (Figure 6). Once all parts are completed, the team discusses ChatGPT's responses and agrees on the final result. This is the simplest form of collaboration, where members work alone on their part and only come together at the end to combine their individual contributions.

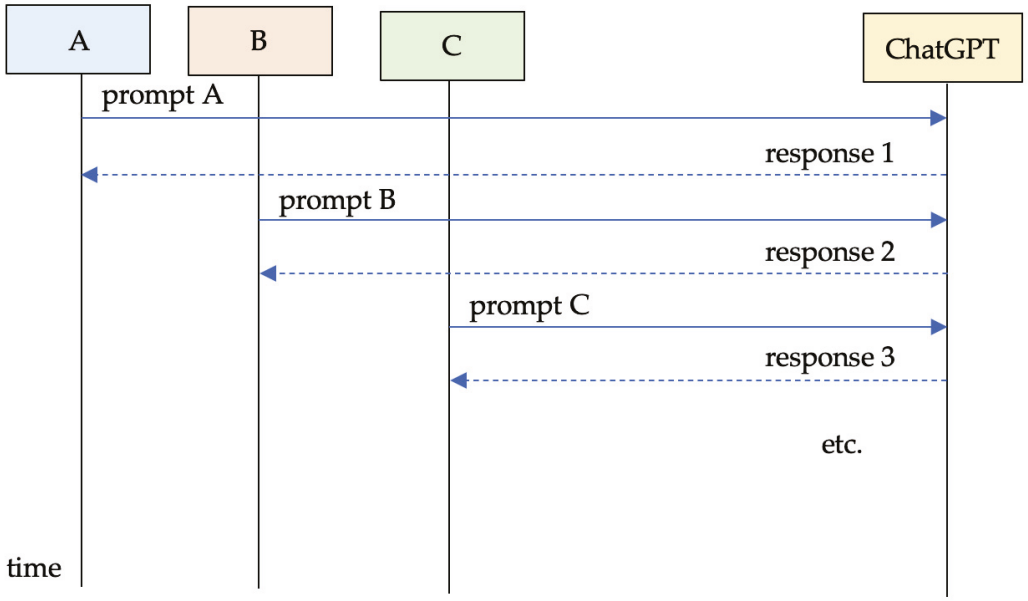


Figure 6. Collaborative *Mode #5* (independently) sequence diagram of team ABC (members: A, B, and C) when interacting with ChatGPT.

4. Quantitative Analysis and Results

This section presents the results of the quantitative analysis of the answers to the closed-ended questions of the questionnaire and the grades of the students' outputs. The statistical analysis of the data included descriptive statistics, Spearman's Rho tests, and Pearson correlations.

4.1. Students Profiles

Out of 256 students who took the course, 51 teams of 3 students (153 students) delivered a report and a presentation. The rest of the students decided to not conduct this work and lost 25% of their final grade. The teams chose to explore the following data mining techniques: classification (16 teams), clustering (15 teams), forecasting (10 teams), sequences (6 teams), associations (3 teams), and anomaly detection (1 team). Then, 10 or 11 teams were randomly assigned to each collaborative mode.

Out of 153 students who delivered a report and a presentation, 136 students answered the questionnaire. Most students (97.1%) were under 22 years old. Almost half of them used the internet for 4–6 h daily (47.8%), a third of them used it for 2–4 h (32.4%), and a tenth of them used it for 4–8 h daily (11.8%). The rest, 8%, used it less than 2 h or more than 8 h daily.

4.2. Students' Preferences Regarding Collaborative Modes

Students possess diverse skills and preferences, with some preferring structured collaborative settings and others favoring flexibility and independence. By experimenting with different collaboration structures, they can develop essential skills such as communication, teamwork, conflict resolution, and leadership. Investigating various collaborative patterns helps educators create effective learning environments that promote inclusivity and en-

gagement. In this study, students did not have strong preferences regarding a specific Collaboration Mode (Table 2, Figure 7). On average, their responses regarding their liking and preference of each Collaboration Mode were close to neutral (M = 3). They marginally liked *Mode #1* (M = 3.12, SD = 1.14), *Mode #2* (M = 3.29, SD = 1.31), *Mode #4* (M = 3.32, SD = 1.12), and *Mode #5* (M = 3.15, SD = 1.29). On the other hand, they marginally did not like *Mode #3* (M = 2.94, SD = 0.99).

Table 2. Students' liking and preference for each Collaboration Mode when interacting with ChatGPT.

Personally, I Like and Prefer this Collaborative Mode (*n* = 136)	Not Used	Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5	Mean	SD
1: One in a row	70	3	19	22	11	11	3.12	1.14
2: Team-oriented	70	6	15	15	14	16	3.29	1.31
3: Member then team	68	5	16	29	14	4	2.94	0.99
4: Team to member	71	5	9	20	22	9	3.32	1.12
5: Independently	63	7	14	18	17	17	3.15	1.29

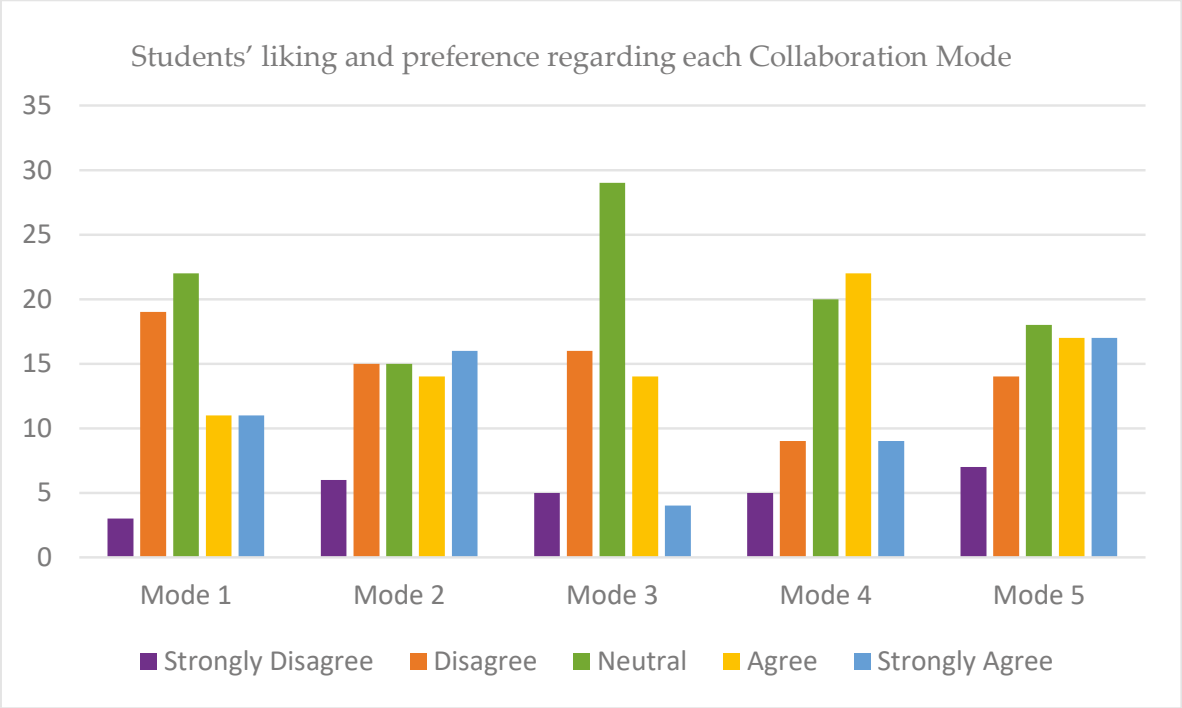


Figure 7. Students' liking and preference for each Collaboration Mode when interacting with ChatGPT (*n* = 136).

More specifically, they agreed and strongly agreed that they liked *Mode #1* (16.2%), *Mode #2* (22.1%), *Mode #3* (13.2%), *Mode #4* (22.8%), and *Mode #5* (25%). So, it seems that they have a slight preference for *Mode #2*, *Mode #4*, and *Mode #5*.

The moderate preferences of students for some collaborative modes implies that students are flexible but potentially uncertain about how collaboration structures affect their learning, indicating a need for clearer scaffolding. Thus, there is a need for adaptability and flexible AI-supported pedagogies, grounded on differentiated instruction in digital education [39].

The slight preference for *Modes #2* and *#5* indicates the contrast between tight collaboration (emphasizing team coherence) and independent work (emphasizing autonomy), as discussed in the collaborative learning literature [16]. The findings show that technology

integration into group work requires collaborative structures that support both individual agency and collective knowledge construction [40].

4.3. Initial Disagreements Among Team Members

During the implementation of their work, teams faced several critical decisions that required thoughtful discussion and deliberation. The team members discussed each issue, presenting their arguments and perspectives, which encouraged critical thinking and collaboration. They worked together to reach consensus and develop their final output. On average, they initially disagreed 30.44% of the time but eventually reached an agreement. More specifically, half of the students (50%) initially disagreed on 20% of the issues, 43 students (31.6%) initially disagreed on 40% of the issues, and 15 students (11%) initially disagreed on 60% of the issues (Figure 8).

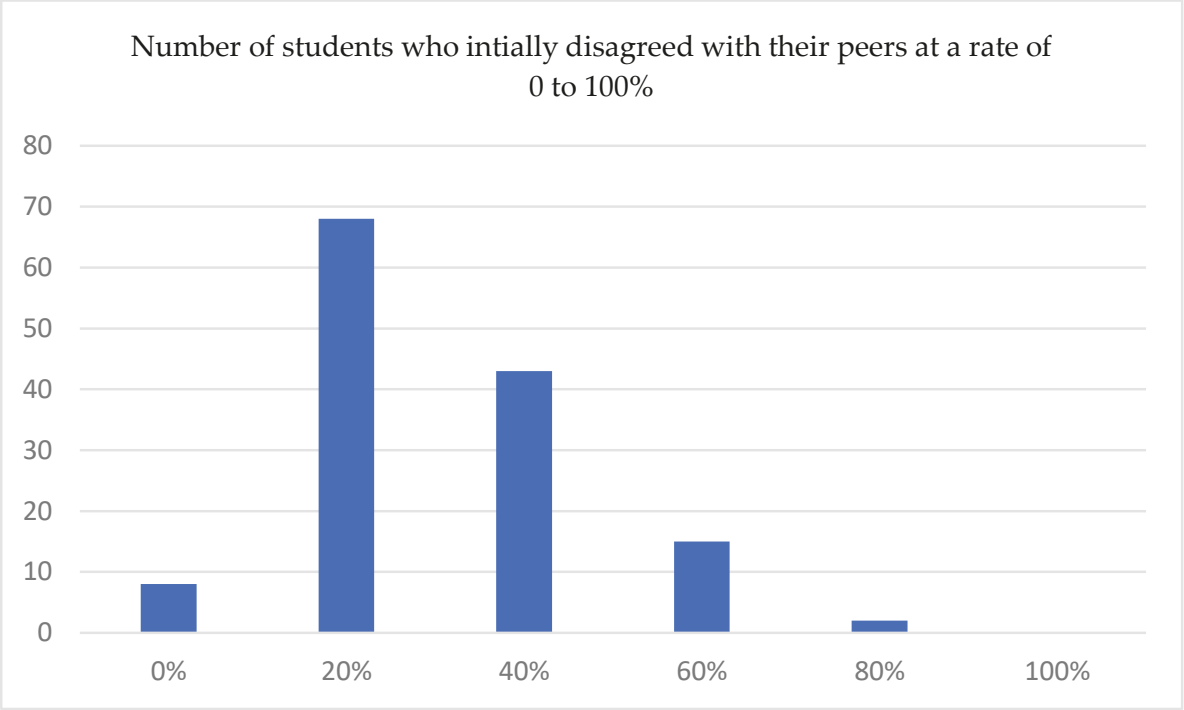


Figure 8. The number of students who initially disagreed with their peers at a rate of 0 to 100% (n = 136).

The initial disagreements and eventual consensus among teammates are related to social constructivism [41], where knowledge is co-constructed through interaction. Note that socio-constructivist learning theory emphasizes the importance of dialogue and negotiation in knowledge construction. The initial disagreements followed by consensus-building are also related to the socio-cognitive conflict process [42], where productive tensions in collaborative learning lead to deeper understanding. The reported disagreements (30.44% initially) align with conflict theory in collaborative learning, where cognitive debate can drive deeper engagement [43].

Also, on average, students input about six prompts into ChatGPT (M = 5.74, SD = 2.81). In fact, only 19.9% of students input more than 10 prompts into ChatGPT. The low use of ChatGPT indicates that students may not fully exploit AI's potential, indicating concerns about superficial engagement [44].

4.4. Extent of Modifying ChatGPT's Responses

Knowing the extent to which students modify ChatGPT's responses helps educators understand how effectively students engage with AI-generated content. It can inform educators whether students use ChatGPT to comprehend the educational subject or whether they just copy its responses. On average, students modified the responses of ChatGPT at a rate of 50% (SD = 0.17). Most students (44.1% of students, n = 60) reported that they modified ChatGPT's responses at a rate of 60% (Figure 9). Fewer students (36% of students, n = 49) reported that they modified ChatGPT's responses at a rate of 40%. Similar fractions of students (11%, n = 15 and 7.4%, n = 10) reported that they modified ChatGPT's responses at a rate of 20% and 80%, correspondingly.

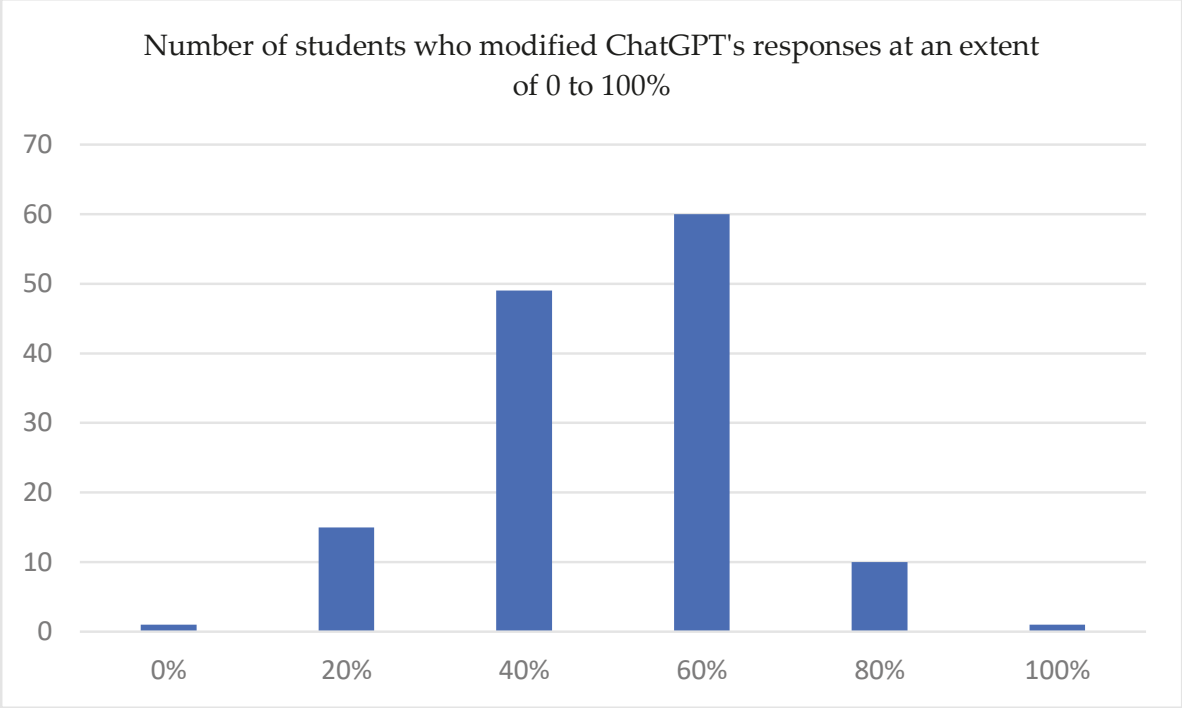


Figure 9. The number of students who modified ChatGPT's responses to an extent of 0% to 100% (n = 136).

The substantial modification rate (50% average) of ChatGPT's responses aligns with Bloom's revised taxonomy [45], implying that students are engaging in higher-order thinking, including analysis, evaluation, and creation, rather than mere knowledge consumption. This finding is supported by constructionist theory [46], where learners actively construct knowledge through manipulating external objects—in this case, AI-generated text.

The fact that students often modified ChatGPT's responses shows their active engagement with AI and supports cognitive apprenticeship models [47], where students refine AI-generated output through their own judgment. This means that students are not passive consumers of AI content but rather active participants in constructing knowledge—supporting the use of AI as a cognitive partner [48].

4.5. Extent of Students' Final Output Based on ChatGPT's Responses

Understanding the percentage of ChatGPT's responses used by students in their final output is essential for assessing academic integrity and the development of critical thinking skills. Determining whether students simply copy responses or significantly modify them reveals their engagement with the material. This finding can inform educators about the effectiveness of AI tools in promoting genuine learning versus dependency on fabricated

material. On average, a fraction, 44% (SD = 0.20), of students' final output was based on ChatGPT's responses. Most students (36% of students, n = 49) reported that their final output was 40% based on ChatGPT's responses (Figure 10). Similar fractions of students (26.5%, n = 36; and 25%, n = 34) reported that their final output was 20% and 60% based on ChatGPT responses. Fewer students (11.8%, n = 16) reported that their final output was 80% based on ChatGPT responses.

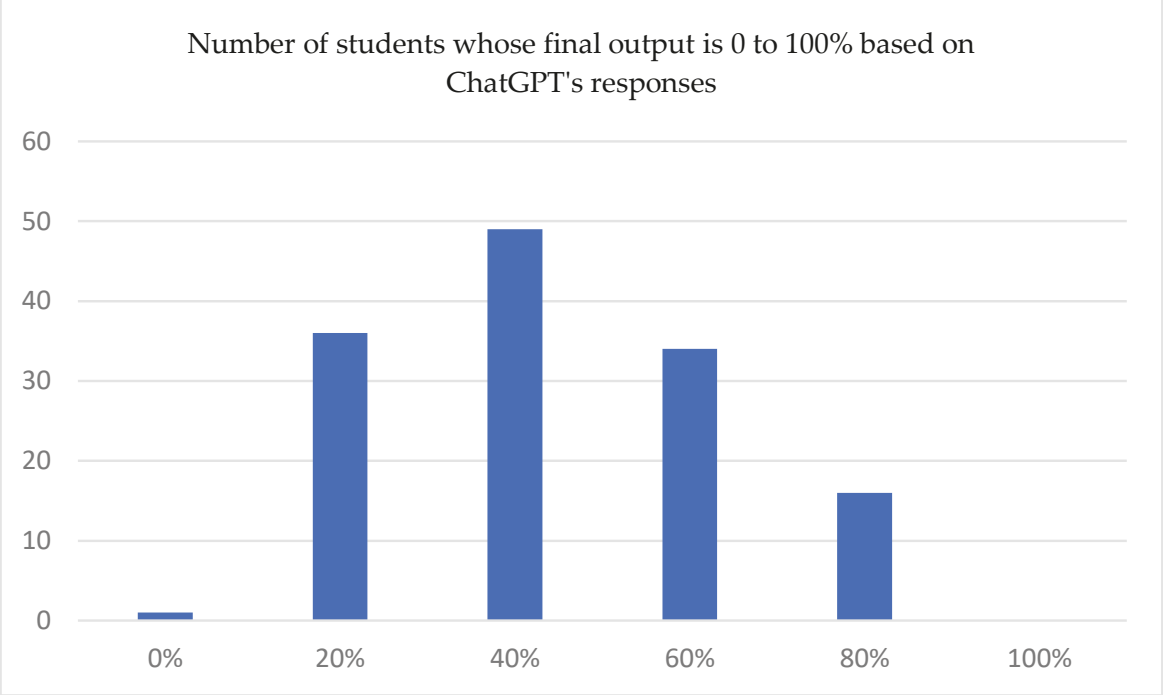


Figure 10. Number of students whose final output is 0 to 100% based on ChatGPT's responses (n = 136).

This finding is related to concerns in AI ethics and academic integrity, questioning the role of AI as a co-author [10]. It shows the contrast between augmentation and overreliance on AI, where educational outcomes depend on students' critical engagement rather than passive acceptance of AI outputs. From a learning perspective, the moderate reliance on AI-generated output shows the importance of improving meta-cognitive skills so that students evaluate and refine AI suggestions rather than overdepend on them [44,49].

The distribution of reliance on AI across different percentages (20–80%) is related to the distributed intelligence framework [50], where cognition is shared between humans and technological artifacts in complementary ways. In addition, the distribution indicates the divergent student approaches of using AI as a cognitive tool that supports and extends but does not replace human cognition and critical thinking in learning environments [51].

4.6. Student Learning During Collaboration

Students can learn significantly during collaborative activities as they engage in discussions, share diverse perspectives, and solve problems together. Students agreed that they learned a lot about their specific data mining technique during their work with their team (M = 4.18, SD = 0.69) (Table 3). Slightly fewer students agreed that their team partners learned a lot about their specific data mining technique during their work with their team (M = 4.00, SD = 0.82). Even fewer students agreed that they learned a lot about other data mining techniques from presentations on the other topics (M = 3.84, SD = 0.90).

Table 3. Student learning during collaboration.

Student Learning During Collaboration (n = 136)	Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5	Mean	SD
While developing the work of our team, I learned a lot about our team's topic.	0	2 (1.5%)	16 (11.8%)	74 (54.4%)	44 (32.4%)	4.18	0.69
While developing the work of our team, I believe that the other members of my team learned a lot about our team's topic.	0	6 (4.4%)	28 (20.6%)	63 (46.3%)	39 (28.7%)	4.00	0.82
I learned a lot from other teams' presentations on their topics.	3 (2.2%)	4 (2.9%)	37 (27.2%)	60 (44.1%)	32 (23.5%)	3.84	0.90

More specifically, 74 students (54.4%) agreed and 44 students (32.4%) strongly agreed that they learned a lot about their specific data mining technique during their team's work (Figure 11). Slightly fewer students, 63 (46.3%), agreed and 39 students (28.7%) strongly agreed that their team partners gained a lot of knowledge about their specific data mining technique during their work. In fact, students believed that they learned slightly more than their team partners. By conducting Spearman's Rho tests, it was found that the association between the grades of the teams' final outputs and their learning perceptions would be considered statistically significant (rs = 0.40393, p (two-tailed) = 0). Also, the association between the grades of the teams' final outputs and their perceptions about their team members' learning would be considered statistically significant (rs = 0.3358, p (two-tailed) = 0.00011).

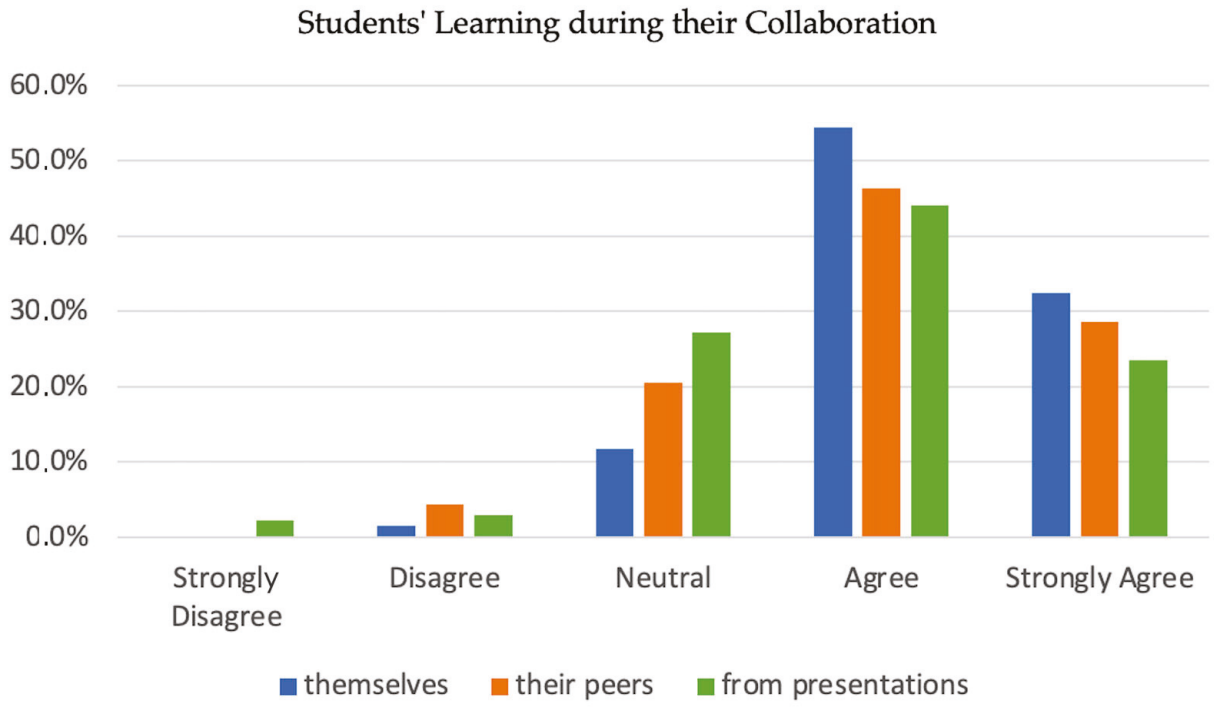


Figure 11. Student perceptions about their own learning and their team members' learning while developing their work and during other teams' presentations (n = 136).

In addition, the association between the students' perceptions that they learnt a lot and also that their team members learnt a lot would be considered statistically significant (Spearman's Rho test: rs = 0.77404, p (two-tailed) = 0; Pearson Correlation Coefficient: r(133) = 0.6954, p = 0.00001).

Finally, their agreement is lower with regard to the extent of their learning about other data mining techniques from the presentations by other teams. A total of 60 students (44.1%) agreed and 32 students (23.5%) strongly agreed that they learned a lot from the presentations of the other teams.

The high self-reported learning (M = 4.18) aligns with social constructivism [41], social learning theory [52], the community of inquiry framework [40], and the conversational framework [53], which emphasize that collaboration can enhance subject understanding and knowledge construction.

5. Qualitative Analysis and Results

This section presents the results of the content analysis [37,38] of the students' answers to the open-ended questions of the questionnaire, focus group discussions, and the teams' reports and presentations (153 students).

5.1. Student Perspectives on Collaborative Mode Advantages and Limitations

For all collaborative modes, there were more positive than negative comments (Table 4). All collaborative modes were praised for the quick generation of results. Collaborative *Mode #2* supported smooth and continuous collaborative processes that enable consistent and coherent final results. On the other hand, Collaborative *Mode #5* emphasized individual work, supporting individual responsibility and flexibility.

Table 4. Student perspectives on the advantages and limitations of the collaborative modes.

Pros/Cons	Attributes	Frequency	Quote Examples
Collaborative Mode #1			
Pros	Good organization; Equal team member work; Improved results through collective effort; Quick generation of results.	23	*It was quite easy and helped to organize the work and deepen the topic as the prompts had the same basis and the members made small modifications to make the result clearer;* *I liked it more that each member did a different wording for each question and in the end, we collaboratively came to a common conclusion and thus in a way verified the validity of the answers of the tool.*
Cons	Time-consuming agreement process; Difficulty in improving some prompts; Risk of repetition.	9	*The only problem was that in some prompts they could not be done further improvements while staying within the basic prompt given by the first member;* *One problem is that it takes more time to agree with the other members of the task.*
Collaborative Mode #2			
Pros	Collective work; Smoother collaboration process; Faster work; Coherent final results.	19	*Each member could individually ask the questions/improvements that it considered useful for our work, but at the same time the result that came out was coherent as it should be;* *helped us make decisions together so that we could work more as a team, while we were operating even faster.*
Cons	Required a common mindset among team members; Not ideal for independent thinkers.	4	*I personally operate more individually in my thinking and with this method we all had to have a common way of thinking.*

Table 4. *Cont.*

Pros/Cons	Attributes	Frequency	Quote Examples
Collaborative Mode #3			
Pros	Quick results; Mix of individual and teamwork; Better results by comparing team members' prompts.	20	*was quite helpful as we were able to compare the prompts of all team members and get as valid a result as possible;* *there was a mix of both individual (personal prompts) and group work (revised prompts) which led to better results;* *there were many prompts so you understood more about the topic*
Cons	Similar prompts leading to repetitive answers; Potential for confusion with multiple prompts.	12	*the prompts of the team members were very similar and the tool often gave the same answers;* *There were many prompts so you understood more about the topic but at the same time it was easy to get confused.*
Collaborative Mode #4			
Pros	Easy organization; Quick results.	21	*helped us in better and easier organization of our team and also in the better collection of information.*
Cons	Similar answers received; Slow work progress due to repetitive responses.	8	*I didn't like receiving similar answers with similar prompts;* *slow work progress because we got similar answers to multiple questions.*
Collaborative Mode #5			
Pros	Quick results; Independent work; Individual responsibility; Easy organization; Critical thinking development.	24	*it was quite effective since each member was responsible for their piece without any confusion;* *be responsible for your own part of the work;* *It helped me to develop my critical thinking as I was asked to search for information about the topic of my group's work on my own and then cross-check it with that of my team members.*
Cons	Lack of direct collaboration; Inconsistency among members; Difficulty in merging individually created work; Insufficient critical discussion among all members.	8	*since we worked quite separately at the beginning, then it was difficult to have a consistency (same fonts, color* etc.*);* *while a negative was the inconsistency in questions asked to the tool;* *this method discouraged direct collaboration as everyone had their own piece and we weren't collectively working out the questions to address;* *I would like to have more collaboration with my team members.*

Each collaborative mode had its own challenges. In *Collaborative Mode #1*, it was difficult to revise prompts, and there was a risk of prompt repetition. Collaborative Mode #2 was not ideal for independent thinkers since it required a common mindset among team members. In *Collaborative Mode #3*, there was a risk of confusion due to similar prompts. In *Collaborative Mode #4*, there was a redundancy in ChatGPT's responses since students were receiving similar responses. In *Collaborative Mode #5*, there were difficulties in integrating students' work due to inconsistent formatting.

Overall, each method had its strengths and weaknesses. *Collaborative Mode #2* appears to be the most balanced approach, allowing for individual contributions while maintaining team coherence. *Collaborative Mode #5* offered the most individual freedom.

The diversity of students' feedback shows that collaboration is complex, and different students may prefer different approaches based on their working styles and personalities and the specific task at hand. These findings align with collaborative learning theory [16] and adaptive learning design [54], showing that student experiences depend on personal preferences, team dynamics, and the structure of collaboration. These findings also align with the community of inquiry framework [40] which describes how different collaborative structures influence social and cognitive presence in technology-enhanced learning environments.

Modes supporting individual autonomy (e.g., *Mode #5*) align with constructivist and connectivism principles, while those enabling tight group work (e.g., *Mode #2*) align with collaborative learning strategies.

The diversity of students' views illustrates that no single AI collaborative mode fits all learners, emphasizing the need for adaptive pedagogy. Multiple options for collaboration in AI-supported higher education should be available to students.

5.2. Student Perspectives on Communication Tools

Students used a variety of communication methods and tools to communicate and collaborate during the development of their team output. They most frequently mentioned the use of phone calls, Instagram, and Google Drive (Table 5). It seems that they combined direct synchronous communication by phone with asynchronous messaging (e.g., Instagram) and file sharing (e.g., Google Drive).

Table 5. Communication Tools used by Students.

Communication Tools	Frequency	Example Quotes
Phone Calls	25	*I liked that on the phone I felt more direct contact than with emails.*
Instagram	24	*We mostly communicated* via *Instagram messages and Instagram video calls.*
Google Drive	21	*We used Google drive and were working on the same document at the same time;* *Our team communicated* via *a shared document on Google Drive to gather and clarify our information and sources.*
Zoom	16	*Our Zoom meetings were very efficient as it was the right thing to do in terms of audio and video capability.*
Emails	14	*The email was very helpful in sending us various links and files useful for our work.*
In-person Communication	12	*As a group we preferred to meet in person to avoid issues such as possible communication difficulties due to lack of signal;* *mainly in person where questions that would be difficult to answer remotely were easily understood.*
Messenger	11	*As a group we used the Messenger app as it was very convenient and the video calling feature through it is easier to use and more familiar to us.*
Discord	7	*I used discord and we didn't have any problems.*
Skype	7	*We used to talk* via *skype and I liked it because even though there were times when we couldn't meet in person there was a way to continue the work.*
FaceTime	3	*FaceTime, we didn't experience any problems.*
Viber	3	*We mainly communicated by phone and* via *Viber.*

Students mainly valued the ease of use of their preferred communication method (Table 6). They liked communications that provide immediate, real-time, and direct interaction, as well as simultaneously enabling document editing and file sharing.

Table 6. Positive aspects of communication methods.

Key Benefits	Frequency	Example Quotes
Ease of use	18	*As a group we used the Messenger app as it was very convenient and the video calling feature through it is easier to use and more familiar to us.*
Immediate communication	6	*I liked that we had quick interaction and almost immediately sent each other what file we needed for any fixes* etc.
Direct interaction enhanced understanding	6	*We had direct communication and there was a better understanding between us.*
Simultaneous editing and sharing	3	*We worked simultaneously and communicated in real time even though we were remote.*

Regarding communication, students mentioned fewer negative issues than positive aspects (Tables 6 and 7). Students mainly faced scheduling difficulties since it was not easy to find common free time due to conflicting individual programs (Table 7). They also

encountered technical issues such as internet connectivity problems, signal variations, and platform limitations. Finally, the lack of direct and in-person communication caused less engagement and more misunderstandings.

Table 7. Challenges faced in communication.

Key Challenges	Frequency	Example Quotes
Scheduling conflicts	8	*It is difficult to coordinate your schedules;* *we had to find time to enter all 3 of us at the same time;* *the difficulty of finding common free time to get the job done;* *The only problem I faced was the difficulty of finding a common time for cooperation.*
Less engaging remote communication and potential misunderstandings	4	*It is quite difficult to collaborate from a distance;* *over the phone there are many times misinterpretation;* *our communication was difficult due to the fact that we were not talking face to face which I would have preferred.*
Technical issues: Internet connectivity problems; Signal variations; Platform-specific limitations (e.g., Zoom time restrictions).	4	*the Zoom had a limited time and was closing;* *difficulties due to lack of signal;* *The main problem was that each of us was in our own city and sometimes one didn't have a good signal;* *We had some technical problems with the internet.*

Finally, students mainly suggested increasing face-to-face collaboration (Table 8) since it enables engagement and mutual understanding. They also suggested resolving technical issues and facilitating scheduling coordination.

Table 8. Suggestions for improvement.

Suggestions	Frequency	Example Quotes
Increase Face-to-Face Collaboration	10	*it would have been better to work together in person.*
Enhance Technical Features	3	*I wish we all had a better signal;* *increase Zoom's free time.*
Enhance Scheduling Coordination	1	*difficulty of finding a common cooperation time. Messenger needs to improve live meetings' coordination.*

The tables above provide a comprehensive overview of students' communication experiences, presenting the tools used, their most appreciated characteristics, challenges encountered, and suggestions for improvement. The analysis revealed a complex landscape of communication and collaboration, with students demonstrating adaptability in using various communication platforms, valuing user-friendliness, flexibility, and immediacy while also recognizing the limitations of remote communication. It would be useful for such communication and collaboration tools to also provide scheduling features as well as efficient integration with other tools. While digital platforms offer convenience, most students still value in-person interactions for more effective teamwork. Many students expressed a preference for face-to-face meetings because they think that they are more effective for understanding, interaction, and team dynamics.

These findings align with "media richness" theory [55] and the "media multiplexity" concept [56], which emphasize the importance of multiple communication channels in digital learning. The preference for hybrid (synchronous/asynchronous) tools relates to the theory of "transactional distance" [57], where different communication tools help overcome spatial and temporal separation in technology-mediated learning.

The desire for in-person interaction indicates the importance of face-to-face communication in digital learning settings [58]. Challenges like scheduling conflicts indicate systemic barriers in

blended learning [59]. The challenges with scheduling and technical issues highlight what [60] termed the "digital disconnect" in educational technology implementation.

In summary, successful team communication and collaboration require a balanced approach, adopting both digital tools and personal interaction.

6. Discussion and Implications

This study investigated how undergraduate students prefer to collaborate and communicate with their teammates to develop their teamwork skills when they prompt ChatGPT. It also examined the initial disagreements in teams, the teams' reliance on AI-generated content, and the students' perceived learning. The findings revealed that the students did not show a strong preference for any one collaborative mode. This implies that students may not have distinct expectations regarding collaboration, possibly because they were not familiar with structured teamwork or they did not understand how different modes influence their learning experience. *Modes #2, #4,* and *#5* received marginally higher preference ratings, possibly because they offered well-defined structures and role clarity. Similar preferences were observed in graduate students' collaboration [32]. Specifically, *Mode #2* facilitated continuous teamwork, while *Mode #5* promoted individual autonomy. Conversely, *Mode #3* received lower preference ratings, possibly because students were uncomfortable with transitions between individual and collective work phases. These preference variations indicate the necessity of providing diverse collaborative structures to accommodate varied student profiles and optimize engagement. These findings are important since they show the need for educators to design collaborative assignments to gradually familiarize students with structured collaboration strategies. Universities and instructional designers can use these findings to create flexible collaborative learning environments that accommodate diverse student needs and preferences. Organizations can also use these findings to enhance team dynamics and optimize project-based work.

Each collaborative mode showed distinct strengths and limitations, indicating the complexity of teamwork and diverse student preferences. *Mode #2*, which emphasized cohesive team-centered collaboration, received higher evaluations, possibly due to its clear structure and coherent teamwork. *Mode #5*, which supported individual accountability and responsibility, received positive evaluations, possibly due to its clear structure and autonomy. Conversely, *Mode #1* presented challenges in prompt refinement and redundancy, *Mode #2* required a unified mindset restraining independent thinking, and *Mode #5* encountered difficulties in combining individually developed work. These differences express the trade-offs between teamwork and individual autonomy, suggesting that no single mode suits all students or tasks. The findings demonstrate the importance of aligning collaborative approaches with learning objectives and student preferences. Educators can use these findings to develop tailored collaborative activities that accommodate diverse learning preferences. For instance, *Mode #2* could be used for tasks requiring cohesion, while *Mode #5* may better serve learning objectives that emphasize individual accountability and specialized experience.

Initial disagreements among team members were reasonably common (30.44% average), possibly due to differences in perspectives and critical thinking processes. Similarly, [32] found that graduate students demonstrated a slightly lower initial disagreement rate (24%). These disagreements may originate from differing student profiles, viewpoints, knowledge, and problem-solving approaches or even varying types of work to be conducted. In other words, students were actively engaged in natural deliberations, analytical thinking, and the evaluation of differing viewpoints, thus increasing their collaborative skills. The final resolution of conflicts demonstrates that students are capable of communicating effectively and building consensus. Initial disagreements do not necessarily lead to

negative outcomes; rather, they can serve as a critical starting point for building a shared understanding and trust within a team [61].

Students modified ChatGPT responses to a fair degree (50% average), showing a moderate degree of critical engagement with AI-generated content. A similar modification rate of ChatGPT's responses (48%) was found by [62]. This intermediate modification rate shows a balance between reliance on ChatGPT and independent critical thinking, possibly due to differences in students' skills, confidence, or willingness to rely on AI assistance. Educators can use these findings to design activities that encourage deeper analysis, originality, and the responsible use of AI, supporting independent thought and creativity.

Students based nearly half (44%) of their final output on ChatGPT's responses, showing moderate reliance on AI-generated content. A similar rate of students' work being based on ChatGPT responses (42%) was found by [62]. However, the significant variability in usage (20–80%) reveals that students use AI tools differently, possibly due to differences in students' confidence, conceptual understanding, or commitment to originality. These findings are important for educators, showing the need to guide students toward responsible AI usage, improve students' authentic learning, and maintain academic integrity. Institutions can apply these findings to inform policies that promote critical engagement with AI while discouraging overdependence on AI-generated content.

Students reported a significant degree of learning regarding their own team's work during collaboration (M = 4.18), with a slightly lower degree of learning with respect to their teammates (M = 4.00) and even less from other teams' presentations (M = 3.84). This result is in line with [32] who found a similar pattern in terms of graduate students' own learning (M = 4.6) and their teammates' learning (M = 4.4) from their team's work, as well as in terms of their own learning from other teams' presentations (M = 3.97). This finding indicates that the degree of learning via active engagement and collaboration is higher than that of learning from watching others' presentations.

In addition, own perceived learning and teammates' learning are significantly associated with each other. This result is in line with [32] who also found that the association between graduate students' perceptions that they learnt a lot and also that their team members learnt a lot from their team's work would be considered statistically significant. Therefore, students believed that both themselves and their teammates benefited from their teamwork.

Furthermore, the statistically significant correlations between teams' grades and their perceived learning indicate that subjective assessments of knowledge acquisition align meaningfully with objective academic performance metrics. On the contrary, ref. [32] found that graduate students' grades and their learning perceptions were not correlated. Thus, further research is needed on this issue. Educators can exploit these findings by creating structured collaborative opportunities that enhance both individual and collective learning within teams and between teams.

Students utilized a diverse mix of communication methods and tools, including synchronous channels (phone calls), asynchronous messaging (Instagram), and collaborative workspaces (Google Drive), to coordinate team activities. This technological diversity highlights the complexity of modern collaboration, where students use a combination of different communication media [63]. Students valued ease of use, immediate real-time interaction, and simultaneous document editing features, but they had concerns related to scheduling conflicts, technical issues (e.g., poor connectivity), and the limitations of remote communication (e.g., reduced engagement and misunderstandings). Similarly, ref. [64] found that substantial variation in team virtuality can negatively impact team performance. Educators can use these findings to recommend tools that combine real-time interaction

with advanced file-sharing capabilities while also encouraging periodic face-to-face meetings when feasible. Additionally, institutions could provide training on effective remote collaboration and ensure reliable technical infrastructure to minimize disruptions.

7. Conclusions, Limitations, and Future Research

This study offers empirical evidence on how undergraduate students collaboratively prompt ChatGPT. It examined five distinct collaborative modes in which teams of three students interacted with ChatGPT to facilitate their teamwork. There were 51 teams randomly assigned to these five collaborative modes. This research explored the advantages and disadvantages of each mode, the communication methods and tools employed, the extent of ChatGPT-generated content incorporated into their final output, and the students' degree of learning. Knowing such information may help educators to appropriately design collaborative assignments, identify and mitigate potential conflicts or inefficiencies within teams, and provide personalized feedback to each student and each team.

Regarding RQ1, students demonstrated no strong preference for specific collaborative modes, suggesting flexibility in their approach to teamwork. However, there was a slight preference for *Mode #2* (cohesive team-oriented collaboration) and *Mode #5* (individual responsibility with final integration), which provided a clearer structure and defined roles. This indicates that students value both unified teamwork and individual autonomy depending on the context. Initial disagreements among team members were common (approximately 30%), implying healthy diversity in perspectives and critical thinking. These disagreements were generally resolved through communication and consensus-building, demonstrating students' developing collaboration skills.

Regarding RQ2, students modified ChatGPT's responses by approximately 50% and incorporated about 44% of ChatGPT's suggestions into their final outputs. This balance suggests a thoughtful integration of AI assistance rather than uncritical acceptance or complete rejection. The significant variability (20–80%) in AI reliance indicates diverse attitudes toward AI tools among students.

Regarding RQ3, students' perceptions were the strongest regarding their own learning during teamwork (M = 4.18), with worse perceptions regarding teammates' learning (M = 4.00) and their own learning from other teams' presentations (M = 3.84). The significant correlation between perceived learning and academic performance validates students' self-assessment as a meaningful indicator of educational outcomes.

Regarding RQ4, communication methods varied, combining synchronous conversation (phone calls), asynchronous messaging (Instagram), and collaborative platforms (Google Drive). Students valued ease of use, real-time interaction, and simultaneous editing capabilities, though they encountered challenges with scheduling conflicts, technical issues, and remote communication limitations.

This study makes several important contributions to understanding AI-supported collaborative learning:

- Framework for AI-Enhanced Collaboration: This study provides students' perspectives on different collaborative modes when working with AI tools, offering educators an adaptable framework for various learning contexts.
- Flexible Collaborative Structures: By demonstrating that students do not exhibit a strong preference for any single collaborative mode—and that each mode offers distinct advantages and limitations—these findings highlight the need for educational designs that accommodate diverse students' profiles and tasks.
- Moderate but Varying AI Reliance: Students' partial dependence on ChatGPT indicates a balanced yet diverse level of reliance. This emphasizes the importance of

guiding students toward the responsible, critical use of AI rather than simple adoption or rejection.

- Linking Perceived Learning to Performance: The correlation between students' perceived learning and actual grades implies that subjective assessments can align with objective measures.
- Communication Strategies for Collaboration: A mix of synchronous and asynchronous tools was common; however, scheduling conflicts, technical limitations, and reduced face-to-face interaction remain concerns. Institutions can optimize technology infrastructure, provide relevant training, and encourage in-person meetings when feasible to enhance teamwork quality.

Despite offering important results, this study has several limitations. First, the relatively small and demographically narrow sample restricts the broader applicability of the results. Larger (e.g., thousands of students), more diverse student profiles (e.g., educational levels, educational subjects, backgrounds, countries) could improve the understanding of the impact of ChatGPT-supported collaboration across different contexts. Second, the reliance on self-reported data introduces the risk of social desirability or recall biases. Third, this study was conducted during a single semester, and it did not investigate possible changes in students' behavior and performance over long time periods.

Since this study investigated a new topic, there are many opportunities for future research. Future studies could examine the impact of different collaborative modes on teams' and students' skills, engagement, motivation, satisfaction, and performance. Also, future research could employ longitudinal designs to examine how students' (or even workers) collaborative preferences, AI utilization, and teamwork skills evolve over time. In parallel, future studies may investigate the long-term effects of using AI on learning and wellbeing. Researchers might also conduct discipline-specific analyses to explore how different fields or course requirements affect AI-supported teamwork. Additionally, exploring instructor perspectives could provide complementary insights into the effective integration of AI tools in collaborative tasks. Cross-cultural comparisons could show how diverse cultural contexts influence both AI usage and teamwork preferences. Furthermore, designing innovative assessments that differentiate individual contributions from collective outcomes and investigating the impact of different collaborative modes on students' learning outcomes, skills, engagement, motivation, and overall satisfaction are attractive directions for further enhancing our understanding of AI's pedagogical potential. Finally, future research may compare the proposed collaborative modes with respect to teams' dynamics, engagement, satisfaction, outputs, performance, etc. Then, appropriate collaborative modes could be suggested for different team and student profiles for different educational contexts.

Author Contributions: The authors contributed equally to this work. Conceptualization, M.P. and A.A.E.; methodology, M.P. and A.A.E.; software, M.P. and A.A.E.; validation, M.P. and A.A.E.; formal analysis, M.P. and A.A.E.; investigation, M.P. and A.A.E.; resources, M.P. and A.A.E.; data curation, M.P. and A.A.E.; writing—original draft preparation, M.P. and A.A.E.; writing—review and editing, M.P. and A.A.E.; visualization, M.P. and A.A.E.; supervision, M.P. and A.A.E.; project administration, M.P. and A.A.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by RESEARCH ETHICS COMMITTEE of the University of Macedonia 5/20-12-2023 2023-12-20.

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

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to thank the anonymous reviewers for their constructive suggestions.

Conflicts of Interest: The authors have no conflicts of interest to disclose. The authors declare that they have no known competing financial interests or personal relationships that could appear to have influenced the work reported in this paper.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence.
CoPIR	Collaborative Preparation, Implementation, Revision.
GenAI	Generative artificial intelligence.
GPT	Generative pre-trained transformer.
LLM	Large language model.
M	Mean.
SD	Standard deviation.

References

1. Lim, W.M.; Gunasekara, A.; Pallant, J.L.; Pallant, J.I.; Pechenkina, E. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *Int. J. Manag. Educ.* **2023**, *21*, 100790. [CrossRef]
2. Kasneci, E.; Seßler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gaseer, U.; Groh, G.; Gunnemann, S.; Hullermeier, E.; et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* **2023**, *103*, 102274. [CrossRef]
3. Rejeb, A.; Rejeb, K.; Appolloni, A.; Treiblmaier, H.; Iranmanesh, M. Exploring the impact of ChatGPT on education: A web mining and machine learning approach. *Int. J. Manag. Educ.* **2024**, *22*, 100932. [CrossRef]
4. Lund, B.D.; Wang, T.; Mannuru, N.R.; Nie, B.; Shimray, S.; Wang, Z. ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *J. Assoc. Inf. Sci. Technol.* **2023**, *74*, 570–581. [CrossRef]
5. OpenAI. ChatGPT [Large Language Model]. 2025. Available online: https://chat.openai.com/chat (accessed on 10 April 2024).
6. Rasul, T.; Nair, S.; Kalendra, D.; Robin, M.; de Oliveira Santini, F.; Ladeira, W.J.; Sun, M.; Day, I.; Rather, R.A.; Heathcote, L. The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *J. Appl. Learn. Teach.* **2023**, *6*, 41–56. [CrossRef]
7. SimilarWeb. Chatgpt.com Website Analysis for February 2025. 2025. Available online: https://www.similarweb.com/website/chatgpt.com/#overview (accessed on 10 March 2025).
8. Montenegro-Rueda, M.; Fernández-Cerero, J.; Fernández-Batanero, J.M.; López-Meneses, E. Impact of the implementation of ChatGPT in education: A systematic review. *Computers* **2023**, *12*, 153. [CrossRef]
9. Crompton, H.; Burke, D. The educational affordances and challenges of ChatGPT: State of the field. *TechTrends* **2024**, *68*, 380–392. [CrossRef]
10. Cotton, D.R.E.; Cotton, P.A.; Shipway, J.R. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innov. Educ. Teach. Int.* **2024**, *61*, 228–239. [CrossRef]
11. Tlili, A.; Shehata, B.; Adarkwah, M.A.; Bozkurt, A.; Hickey, D.T.; Huang, R.; Agyemang, B. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learn. Environ.* **2023**, *10*, 15. [CrossRef]
12. Tan, S.C.; Lee, A.V.Y.; Lee, M. A systematic review of artificial intelligence techniques for collaborative learning over the past two decades. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100097. [CrossRef]
13. Thorp, H.H. ChatGPT is fun, but not an author. *Science* **2023**, *379*, 313. [CrossRef]
14. Chen, J.; Wang, M.; Kirschner, P.A.; Tsai, C.C. The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Rev. Educ. Res.* **2018**, *88*, 799–843. [CrossRef]
15. Salomon, G.; Perkins, D.N. Individual and social aspects of learning. *Rev. Res. Educ.* **1998**, *23*, 1–24. [CrossRef]
16. Dillenbourg, P. What do you mean by collaborative learning. In *Collaborative-Learning: Cognitive and Computational Approaches*; Dillenbourg, P., Ed.; Elsevier: Oxford, UK, 1999; pp. 1–19.
17. Prince, M. Does active learning work? A review of the research. *J. Eng. Educ.* **2004**, *93*, 223–231. [CrossRef]

18. Springer, L.; Stanne, M.E.; Donovan, S.S. Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Rev. Educ. Res.* **1999**, *69*, 21–51. Available online: http://www.jstor.org/stable/1170643 (accessed on 10 April 2024). [CrossRef]
19. Gilson, A.; Safranek, C.W.; Huang, T.; Socrates, V.; Chi, L.; Taylor, R.A.; Chartash, D. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR Med. Educ.* **2023**, *9*, e45312. [CrossRef]
20. Mena-Guacas, A.F.; Urueña Rodríguez, J.A.; Santana Trujillo, D.M.; Gómez-Galán, J.; López-Meneses, E. Collaborative learning and skill development for educational growth of artificial intelligence: A systematic review. *Contemp. Educ. Technol.* **2023**, *15*, ep428. [CrossRef]
21. Kumar, J.A. Educational chatbots for project-based learning: Investigating learning outcomes for a team-based design course. *Int. J. Educ. Technol. High. Educ.* **2021**, *18*, 65. [CrossRef]
22. Sikström, P.; Valentini, C.; Sivunen, A.; Kärkkäinen, T. How pedagogical agents communicate with students: A two-phase systematic review. *Comput. Educ.* **2022**, *188*, 104564. [CrossRef]
23. Haq, I.U.; Anwar, A.; Rehman, I.U.; Asif, W.; Sobnath, D.; Sherazi, H.H.R. Dynamic group formation with intelligent tutor collaborative learning: A novel approach for next generation collaboration. *IEEE Access* **2021**, *9*, 143406–143422. [CrossRef]
24. Wiboolyasarin, W.; Wiboolyasarin, K.; Suwanwihok, K.; Jinowat, N.; Muenjanchoey, R. Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Comput. Educ. Artif. Intell.* **2024**, *6*, 100228. [CrossRef]
25. Msambwa, M.M.; Wen, Z.; Daniel, K. The impact of AI on the personal and collaborative learning environments in higher education. *Eur. J. Educ.* **2025**, *60*, e12909. [CrossRef]
26. Kovari, A. A systematic review of AI-powered collaborative learning in higher education: Trends and outcomes from the last decade. *Soc. Sci. Humanit. Open* **2025**, *11*, 101335. [CrossRef]
27. Ouyang, F.; Zhang, L. AI-driven learning analytics applications and tools in computer-supported collaborative learning: A systematic review. *Educ. Res. Rev.* **2024**, *44*, 100616. [CrossRef]
28. Kong, X.; Liu, Z.; Chen, C.; Liu, S.; Xu, Z.; Tang, Q. Exploratory study of an AI-supported discussion representational tool for online collaborative learning in a Chinese university. *Internet High. Educ.* **2025**, *64*, 100973. [CrossRef]
29. Darmawansah, D.; Rachman, D.; Febiyani, F.; Hwang, G.J. ChatGPT-supported collaborative argumentation: Integrating collaboration script and argument mapping to enhance EFL students' argumentation skills. *Educ. Inf. Technol.* **2025**, *30*, 3803–3827. [CrossRef]
30. Zheng, L.; Fan, Y.; Gao, L.; Huang, Z.; Chen, B.; Long, M. Using AI-empowered assessments and personalized recommendations to promote online collaborative learning performance. *J. Res. Technol. Educ.* **2024**, 1–27. [CrossRef]
31. Albadarin, Y.; Saqr, M.; Pope, N.; Tukiainen, M. A systematic literature review of empirical research on ChatGPT in education. *Discov. Educ.* **2024**, *3*, 60. [CrossRef]
32. Perifanou, M.; Economides, A.A. Collaborative uses of GenAI tools in project-based learning. *Educ. Sci.* **2025**, *15*, 354. [CrossRef]
33. Fusch, P.; Fusch, G.E.; Ness, L.R. Denzin's paradigm shift: Revisiting triangulation in qualitative research. *J. Sustain. Soc. Change* **2018**, *10*, 19–32. [CrossRef]
34. Thurmond, V.A. The point of triangulation. *J. Nurs. Scholarsh.* **2001**, *33*, 253–258. [CrossRef]
35. Carter, N. The use of triangulation in qualitative research. *Oncol. Nurs. Forum* **2014**, *41*, 545–547. [CrossRef]
36. Johnson, R.B.; Onwuegbuzie, A.J.; Turner, L.A. Toward a definition of mixed methods research. *J. Mix. Methods Res.* **2007**, *1*, 112–133. [CrossRef]
37. Hsieh, H.F.; Shannon, S.E. Three approaches to qualitative content analysis. *Qual. Health Res.* **2005**, *15*, 1277–1288. [CrossRef]
38. Neuendorf, K.A. Content analysis and thematic analysis. In *Advanced Research Methods for Applied Psychology*; Routledge: London, UK, 2018; pp. 211–223.
39. Ziernwald, L.; Hillmayr, D.; Holzberger, D. Promoting high-achieving students through differentiated instruction in mixed-ability classrooms—A systematic review. *J. Adv. Acad.* **2022**, *33*, 540–573. [CrossRef]
40. Garrison, D.R. *E-learning in the 21st Century: A Community of Inquiry Framework for Research and Practice*, 3rd ed.; Routledge: London, UK, 2016. [CrossRef]
41. Vygotsky, L.S. *Mind in Society: The Development of Higher Psychological Processes*; Harvard University Press: Cambridge, MA, USA, 1978; Volume 86.
42. Doise, W.; Mugny, G. *The Social Development of the Intellect*; Oxford University Press: Oxford, UK, 1984.
43. Jehn, K.A.; Mannix, E.A. The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Acad. Manag. J.* **2001**, *44*, 238–251. [CrossRef]
44. Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. Systematic review of research on artificial intelligence applications in higher education–where are the educators? *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 39. [CrossRef]
45. Krathwohl, D.R. A revision of Bloom's taxonomy: An overview. *Theory Into Pract.* **2002**, *41*, 212–218. [CrossRef]

46. Papert, S.A. *Mindstorms: Children, Computers, and Powerful Ideas*; Basic books: New York, NY, USA, 2020.
47. Collins, A.; Brown, J.S.; Newman, S.E. Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In *Knowing, Learning, and Instruction*; Routledge: London, UK, 1989; pp. 453–494. [CrossRef]
48. Pea, R.D. The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *J. Learn. Sci.* **2004**, *13*, 423–451. [CrossRef]
49. Wijaya, T.T.; Yu, Q.; Cao, Y.; He, Y.; Leung, F.K. Latent Profile Analysis of AI Literacy and Trust in Mathematics Teachers and Their Relations with AI Dependency and 21st-Century Skills. *Behav. Sci.* **2024**, *14*, 1008. [CrossRef]
50. Pea, R.D. Practices of distributed intelligence and designs for education. *Distrib. Cogn. Psychol. Educ. Consid.* **1993**, *11*, 47–87.
51. Jonassen, D.H. *Computers as Mindtools for Schools: Engaging Critical Thinking*, 2nd ed.; Prentice Hall: Saddle River, NJ, USA, 2000. Available online: https://lccn.loc.gov/99032758 (accessed on 31 March 2025).
52. Bandura, A.; Walters, R.H. *Social Learning Theory*; Prentice hall: Englewood Cliffs, NJ, USA, 1977; Volume 1, pp. 141–154.
53. Laurillard, D. *Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies*, 2nd ed.; Routledge: London, UK, 2013. [CrossRef]
54. Economides, A.A. Adaptive context-aware pervasive and ubiquitous learning. *Int. J. Technol. Enhanc. Learn.* **2009**, *1*, 169–192. [CrossRef]
55. Daft, R.L.; Lengel, R.H. Organizational information requirements, media richness and structural design. *Manag. Sci.* **1986**, *32*, 554–571. [CrossRef]
56. Haythornthwaite, C. Social networks and Internet connectivity effects. *Inf. Community Soc.* **2005**, *8*, 125–147. [CrossRef]
57. Garrison, R. Theoretical challenges for distance education in the 21st century: A shift from structural to transactional issues. *Int. Rev. Res. Open Distrib. Learn.* **2000**, *1*, 1–17. [CrossRef]
58. Bali, S.; Liu, M.C. Students' perceptions toward online learning and face-to-face learning courses. *J. Phys. Conf. Ser.* **2018**, *1108*, 012094. [CrossRef]
59. Graham, C.R. Emerging practice and research in blended learning. In *Handbook of Distance Education*; Moore, M.G., Ed.; Routledge: London, UK, 2013; pp. 333–350.
60. Selwyn, N. Exploring the 'digital disconnect' between net-savvy students and their schools. *Learn. Media Technol.* **2006**, *31*, 5–17. [CrossRef]
61. Gevers, J.; Li, J.; Rutte, C.G.; Eerde, W.v. How dynamics in perceptual shared cognition and team potency predict team performance. *J. Occup. Organ. Psychol.* **2019**, *93*, 134–157. [CrossRef]
62. Economides, A.A.; Perifanou, M. Higher education students using GenAI tools to design websites. In Proceedings of the Innovating Higher Education conference (I-HE2024) Proceedings, Limassol, Cyprus, 23–25 October 2024. [CrossRef]
63. Kaneko, A. Team Communication in the workplace: Interplay of communication channels and performance. *Bus. Prof. Commun. Q.* **2023**, *57*, 370–400. [CrossRef]
64. Hincapie, M.X.; Hill, N.S. The impact of team virtuality on the performance of on-campus student teams. *Acad. Manag. Learn. Educ.* **2024**, *23*, 158–175. [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.

computers

MDPI

Systematic Review

Artificial Intelligence in K-12 Education: A Systematic Review of Teachers' Professional Development Needs for AI Integration

Spyridon Aravantinos [1], Konstantinos Lavidas [1], Vassilis Komis [1], Thanassis Karalis [1] and Stamatios Papadakis [2,*]

[1] Department of Educational Sciences and Early Childhood Education, University of Patras Greece, 26504 Patras, Greece; aravantinos_spyridon@upatras.gr (S.A.); lavidas@upatras.gr (K.L.); komis@upatras.gr (V.K.); karalis@upatras.gr (T.K.)
[2] Department of Preschool Education, University of Crete, 74100 Rethymno, Greece
* Correspondence: stpapadakis@uoc.gr

Abstract

Artificial intelligence (AI) is reshaping how learning environments are designed and experienced, offering new possibilities for personalization, creativity, and immersive engagement. This systematic review synthesizes 43 empirical studies (Scopus, Web of Science) to examine the training needs and practices of primary and secondary education teachers for effective AI integration and overall professional development (PD). Following PRISMA guidelines, the review gathers teachers' needs and practices related to AI integration, identifying key themes including training practices, teachers' perceptions and attitudes, ongoing PD programs, multi-level support, AI literacy, and ethical and responsible use. The findings show that technical training alone is not sufficient, and that successful integration of AI requires a combination of pedagogical knowledge, positive attitudes, organizational support, and continuous training. Based on empirical data, a four-level, process-oriented PD framework is proposed, which bridges research with educational practice and offers practical guidance for the design of AI training interventions. Limitations and future research are discussed.

Keywords: AI; primary education; secondary education; K-12 education; teacher training; AI literacy; transformative learning

1. Introduction

Artificial intelligence (AI) is already present in educational environments and is influencing the educational landscape, placing teachers at the center of these developments. However, the pivotal role of teachers as mediators and facilitators of interactions between students and AI tools as not been sufficiently explored [1,2]. Digital literacy in AI is essential for teachers to effectively integrate it into their teaching practices [3], and professional development (PD) programs should explicitly link learning theories with the practical implementation of AI to successfully support this integration [4]. AI literacy, which is also important for students' learning, can enhance teachers' ability to incorporate AI technologies into the learning process, and its development has become a priority in many countries [5], following directives from international organizations, such as AI competency frameworks for teachers and students from United Nations Educational, Scientific and Cultural Organization (UNESCO) [6]. Building on the conceptualization of knowledge necessary for integrating emerging technologies into education [7], the Technological Pedagogical Content Knowledge (TPACK) framework underscores the dynamic interaction among technology, pedagogy, and content knowledge, which is critical for effective tech-

nology adoption [8]. Recent extensions of this framework, such as Intelligent-TPACK [9] and AI-TPACK [10], aim to align teacher competencies with the evolving demands of AI.

Nowadays, teachers are called to participate in AI PD programs, the majority of which focus on the utilization of continuously emerging AI applications. Such PD programs emerge in rapid succession, a phenomenon attributable both to economic incentives and to the growing demand for AI-based tools. Practical training that lacks pedagogical orientation does not ensure that teachers will ultimately integrate AI into classroom practice. Such approaches, unfortunately, do little to support the effective integration of AI into pedagogical design, nor do they contribute to the development of teachers' pedagogical competence in relation to AI [11]. Moreover, to the best of our knowledge, previous systematic reviews do not synthesize empirical evidence for tangible and immediately implementable AI-focused teacher PD. To cover this gap, this systematic review aims to investigate teachers' views of AI and its implementation in teaching and learning processes by documenting their training needs as shaped by the rapidly evolving technological landscape. Furthermore, the collection of empirical evidence from the papers analyzed in this systematic review will inform the development of a process-oriented PD framework.

This systematic literature review investigates teachers' training needs for AI use and corresponding PD in primary and secondary education. Training needs are conceptualized as the knowledge, skills, and support required for the pedagogically effective and ethical integration of AI in classroom practice. The focus on primary and secondary education reflects the developmental characteristics of these levels, where learning is structured and AI use is predominantly teacher-driven [12].

The findings of this study could serve the best interests of multiple stakeholders [13]: teachers, who strengthen their professional resilience; school leaders, who enhance institutional performance; educational consultants, who organize training programs supporting teachers; and students, who benefit from AI-literate teachers capable of personalizing learning. The necessity of continuous PD for teachers across diverse domains, including AI integration, is well established. Consequently, AI competence should be embedded within PD programs by policymakers and educational leaders through the design of training initiatives that strengthen teachers' knowledge and skills for the effective use of AI in school settings [14].

2. Previous Systematic Literature Reviews

In Table A3, there is a list of previous systematic reviews referenced to in this section, along with their basic findings. Ref. [15] systematically reviewed 67 articles related to AI-powered chatbots in education and highlighted enhancements in personalization and feedback, but also the need for PD training for teachers and the demand to implement measures for ethical use and the promotion of awareness among students. ChatGPT's use in K-12 education was also investigated by ref. [16], who synthesized 13 articles and performed strengths, weaknesses, opportunities, and threats (SWOT) analysis, underscoring the importance of thoughtful integration and highlighting the practices of teachers. Ethical and regulatory concerns of Generative Artificial Intelligence (GenAI) in education were the focus of a systematic review of 53 peer-reviewed articles, which underlined that addressing these challenges can ensure responsible AI use [17]. Additionally, on the use of GenAI in K-12 education, ref. [18] conducted a systematic review with 197 studies, highlighting gaps in teachers' PD and demonstrating the potential of AI to transform educational systems, while noting that dependency and ethical concerns need to be addressed. It also stated that teacher training, both pre-service and ongoing, must be collaborative, innovative, inclusive, and transparent to ensure the effectiveness of AI-enhanced education and to improve AI literacy among students [18].

https://doi.org/10.3390/computers15010049

Another aspect of teacher education, human–machine dialogic learning, was the focus of ref. [19] in their systematic review of 22 studies. They revealed three core elements: pedagogy, technological tools, and learning environment, three types of affordances: teaching skill and practice, learning experience and participation, and teaching interaction and communication, and four key aspects: dialogic purpose, dialogic medium, dialogic mode, and dialogic application [19]. This review revolved around pre-service teacher education, but dialogic teaching aids can also be a critical tool for in-service teacher PD, allowing them to engage in collaborative conversations and self-reflection, enhance professional skills, and contribute to their professional growth and development [19]. A systematic literature review of 24 papers by ref. [20] provided guidance for the application of human-centered AI (HCAI) in policy formulation, teacher training, and the promotion of technology applications, emphasizing data privacy and ethical issues, algorithmic bias, and guidance for teachers to design personalized instruction with ongoing training at a practical level. In addition, ref. [21] conducted a systematic review of 40 empirical studies unpacking key characteristics of human-centered AI and identifying essential stakeholders, underscoring the necessity of collaborative efforts to ensure that AI does not dehumanize learning and has sustainable and positive outcomes in education.

Regarding AI literacy, ref. [22], in a systematic review of 25 empirical studies, emphasized the critical construct of AI ethics in AI literacy and underscored the need for more diverse and innovative pedagogies, and the development of an inclusive AI literacy competency framework. Key themes identified in diverse conceptualizations of AI literacy include the development of curricula and literacy frameworks tailored to different educational levels [22]. As highlighted by ref. [23] in a systematic review of 87 articles, the integration of AI literacy in K-12 education emerges as a critical priority—one that must cultivate not only technical competencies but also address the ethical dimensions of AI use. A systematic review of 22 studies on AI literacy in K-12 education underscored its absence in the context of teacher education and the research gap for factors influencing AI literacy among students and teachers, for whom AI literacy constitutes a core element of AI-enhanced learning [24]. Evidence from in-service teachers indicates generally positive attitudes toward AI and given its critical role in facilitating the integration of AI-driven applications, AI literacy should be prioritized within PD initiatives [24].

As for the establishment of clear guidelines and regulatory frameworks, it has been expressed that integrated mechanisms for ongoing assessment are required to balance technological progress and individual rights, through the collaboration of all stakeholders: policymakers, teachers, and technologists [17]. Ref. [25] reviewed and classified AI competence frameworks for teachers, emphatically stating the need for both theoretical policy and teacher-driven approaches, while noting challenges such as the avoidance of redundancy with digital competency models. Furthermore, a systematic review of 14 publications on AI and PD emphasized significant gaps in AI PD programs regarding AI tool integration, underscoring the need to promote teachers' teaching skills with AI rather than limiting PD programs to AI content alone [26]. It also highlighted the significance of a guiding framework in AI PD, the need to balance pedagogical, content, and technological knowledge, and the gap in addressing ethical dimensions of AI technologies in teacher PD [26]. AI PD features, formats, and active learning approaches should be further explored, while structured feedback and support mechanisms, practical application with specific AI-based tools, balance between technical expertise and pedagogical effectiveness, incorporation of acquired skills and knowledge in daily teaching routines, and the establishment of an evidence-based conceptual framework have a great impact on teacher outcomes, can inform stakeholders in decision-making, and are crucial for the design of AI PD [26].

https://doi.org/10.3390/computers15010049

All previous systematic reviews agree on the benefits of AI [27] and GenAI [28] such as personalization, accessibility, teaching and learning support, improved learning outcomes, flexibility, and enhanced interaction. They also recognize ethical challenges [27,28] like algorithmic bias, academic integrity, privacy, digital divide, lack of emotional intelligence, and dependence on technology. Some have proposed human-centered approaches to AI to promote equality and social justice through personalization and call for the integration of AI into curricula, the development of guidelines, and AI literacy. The emphasis is placed on pedagogical dimensions, the obligation for cooperation among the stakeholders involved, and the need for continuous teacher PD. Although they recognize the challenges of AI integration and the need for teacher PD, all these systematic reviews are limited in synthesizing empirical evidence for tangible and immediately implementable guidelines.

Objective of This Study

According to the previous discussion, the objective of this study is to systematically review empirical studies to identify the training needs and practices of primary and secondary education teachers for efficient AI integration and overall PD.

3. Methods

This review adopts a qualitative thematic synthesis rather than an effectiveness-focused meta-analysis. To maintain the integrity of this systematic literature review, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Figure 1) [29] were followed. During identification, two of the most robust and rigorously validated databases of research—Scopus and Web of Science—both of which cover only peer-reviewed journals and provide extensive indexing across many disciplines, were used. Because each database includes unique journal content, the utilization of both reduced the potential to omit vital studies. Additionally, their combined search capability enables reproducible queries as per PRISMA standards, thereby enhancing methodological quality. Both databases offer comprehensive coverage within the topic area specific to this review and are globally recognized for containing high-quality scholarly content, thereby capturing a broad body of literature. Table 1 shows the search string queries performed across all databases in February 2025, targeting online literature produced up to the end of the last day of December 2024. A few articles were published in January 2025 but were already online in late December 2024 and were therefore included. The findings of previous systematic reviews focused on the positive aspects of using AI in the educational process [15,16,19], but they also highlighted many challenges [17,22,23], with most of them emphasizing the significant gap in teacher training, which could mitigate these difficulties [18,20,24,26]. As a result, the search terms were set after screening the relevant systematic reviews and were similar or synonymous with “teacher training”, “professional development”, and terms related to the teachers’ educational levels that were the focus of our study.

During screening, English peer-reviewed articles with empirical data and explicit discussion of teachers and AI integration in K–12 classrooms were included. To assess eligibility, articles were screened based on pre-specified inclusion and exclusion criteria (Table 2), first by title and abstract and, if necessary, by full text. Materials excluded consisted of books, conference proceedings, systematic reviews, meta-analyses, and theoretical articles without empirical evidence, as well as studies published in languages other than English, articles in disciplines unrelated to the subject area (e.g., Pharmacology, Immunology, Architecture, Law, Tourism, Ergonomics, and Business) and duplicate documents.

https://doi.org/10.3390/computers15010049

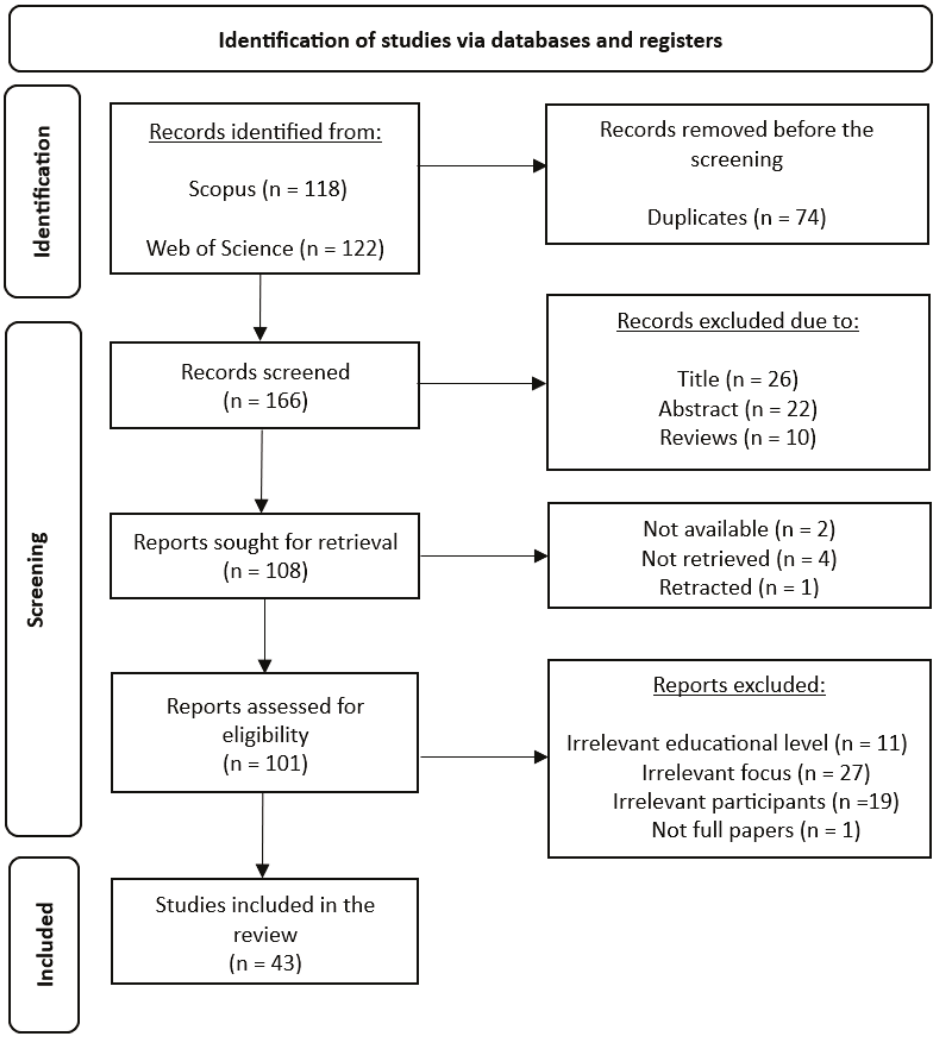


Figure 1. Selection of empirical studies process.

Table 1. Scopus and Web of Science search strings.

Scopus search string
PUBYEAR bef 2025 (TITLE-ABS-KEY(“artificial intelligence” AND (“primary school” OR “elementary school” OR “primary education” OR “elementary education” OR “secondary school” OR “high school” OR “middle school” OR “upper school” OR “secondary education” OR “secondary level” OR educator OR “K-12” OR k12) AND (“professional development” OR “teacher training” OR “continuous education” OR “professional learning” OR “lifelong learning”)) AND TITLE-ABS-KEY(“empirical study” OR “case study” OR survey OR interview OR “data collection” OR “qualitative research” OR “quantitative research” OR “mixed methods” OR “experimental research”))
Web of Science search string Timespan: 1 January 1970 to 31 December 2024 (Publication Date)
TS = (“artificial intelligence” AND (“primary school” OR “elementary school” OR “primary education” OR “elementary education” OR “secondary school” OR “high school” OR “middle school” OR “upper school” OR “secondary education” OR “secondary level” OR educator OR “K-12” OR k12) AND (“professional development” OR “teacher training” OR “continuous education” OR “professional learning” OR “lifelong learning”)) **Combined with:** TS = (“empirical study” OR “case study” OR survey OR interview OR “data collection” OR “qualitative research” OR “quantitative research” OR “mixed methods” OR “experimental research”)

https://doi.org/10.3390/computers15010049

Table 2. Inclusion and exclusion criteria for Scopus- and Web of Science-indexed articles.

Inclusion Criteria	Exclusion Criteria
Empirical studies (peer-reviewed articles) published online up to December 2024	Systematic reviews and meta-analyses, conference papers, and books
Studies focusing on training regarding AI implementation for K-12 teachers	Studies about teaching the subject of AI (except if teaching about AI occurred with the use of AI or if a study referred to both)
Sample: teachers of primary and secondary education (studies with mixed participants were included only if their results were categorized accordingly)	Sample: only students and/or parents
Scientific fields: Arts and Humanities, Computer Science, Mathematics, Multidisciplinary, Neuroscience, Psychology, and Social Sciences	Irrelevant scientific fields: Agricultural and Biological Sciences, Biochemistry, Genetics and Molecular Biology, Business, Management and Accounting, Chemical Engineering, Chemistry, Decision Sciences, Dentistry, Earth and Planetary Sciences, Economics, Econometrics and Finance, Energy, Engineering, Environmental Science, Health Professions, Immunology and Microbiology, Materials Science, Medicine, Nursing, Pharmacology, Toxicology and Pharmaceutics, Physics and Astronomy, and Veterinary
Written in English	All other languages

The search in the databases yielded 240 papers, of which 74 were duplicates, 26 were excluded by title, 22 by abstract, and 10 were reviews of all kinds (systematic, literature, scoping, bibliometric analyses, meta-analyses, and theoretical). Of the remaining 108 papers, 2 were unavailable, 4 could not be retrieved, and 1 was retracted. In the last stage of screening, we assessed for eligibility 101 articles of which 11 were excluded due to different educational level, 27 due to a different focus, 19 did not have eligible participants, and 1 was an incomplete paper, leading to the final 43 included articles (Table A1).

For the purposes of data analysis, a spreadsheet was designed for the systematic organization of important findings guided by the research question of this study. The spreadsheet included identification numbers for all articles and important bibliographic and methodological data such as author(s), title, year of publication, DOI, abstract, keywords, aims of the studies, methodological design and procedures, sample of participants, country of origin, and important findings. The study utilized a reflexive thematic analysis [30]. In this analysis, two researchers independently conducted close readings of the included studies and systematically extracted and coded findings using a shared data extraction spreadsheet. This initial phase focused on data familiarization and inductive code generation. Inter-rater reliability was assessed at this stage and indicated high agreement (85%), which corresponds to a substantial level of reliability (Cohen's $\kappa \approx 0.8$). Divergencies in coding were addressed through review by a third researcher, whose contribution supported reflexive dialog and further code refinement. The research team then engaged in an iterative analytic process, examining codes and data to identify patterns of meaning relevant to the study's focus, particularly teacher training needs and practices. The final codes were organized into broader analytic categories, leading to the inductive development of themes grounded in the empirical evidence.

Descriptives of the Studies

The year of publication of the empirical studies was as follows: 2025 (n = 6), 2024 (n = 31), 2023 (n = 3), 2022 (n = 2), and 2021 (n = 1). Most research took place in China (n = 9), the USA

https://doi.org/10.3390/computers15010049

(*n* = 6), Turkey (*n* = 5), Indonesia (*n* = 2), the Philippines (*n* = 2), Republic of Korea (*n* = 3), Vietnam (*n* = 2), Saudi Arabia (*n* = 2), and in the following countries (*n* = 1): Brazil, Cyprus, Estonia, Hong Kong, Iran, Israel, Japan, Kazakhstan, Kenya, Nepal, Nigeria, Norway, the Russian Federation, Sweden, Taiwan (Republic of China), and Ukraine. Asia was the continent with the most research (*n* = 32), followed by Europe (*n* = 7), North America (*n* = 7), Africa (*n* = 2), Oceania (*n* = 1), and South America (*n* = 1). The research methods were qualitative (*n* = 15), quantitative (*n* = 14), and mixed (*n* = 14), and the participant sample size across the 43 final articles exceeded 7500 in-service teachers.

More specifically, according to whether researchers mentioned educational level or the way they categorized school ages in relevance to each country's system, the number of teachers was as follows: preschool (*n* = 11), primary (*n* = 1.638), secondary (*n* = 4.240), mixed (*n* = 595), K-12 (*n* = 386), higher education (*n* = 264), and N/A (*n* = 297). Since this review focuses only on primary and secondary education, results for preschool and higher education levels were not taken into consideration; however, the articles were included because they also contained participants at the levels of interest. In some studies, parents and students were also participants, but their results were not mentioned. Whenever there was a distinction between teachers and school principals, administrators, or curriculum leaders (*n* = 213), their opinions were recorded, as they were also teachers.

4. Results

Figure 2 presents the contribution of themes indicated in this systematic review. Table A2 in the Appendix A presents all themes with their corresponding codes. The next section analytically presents these themes and their codes.

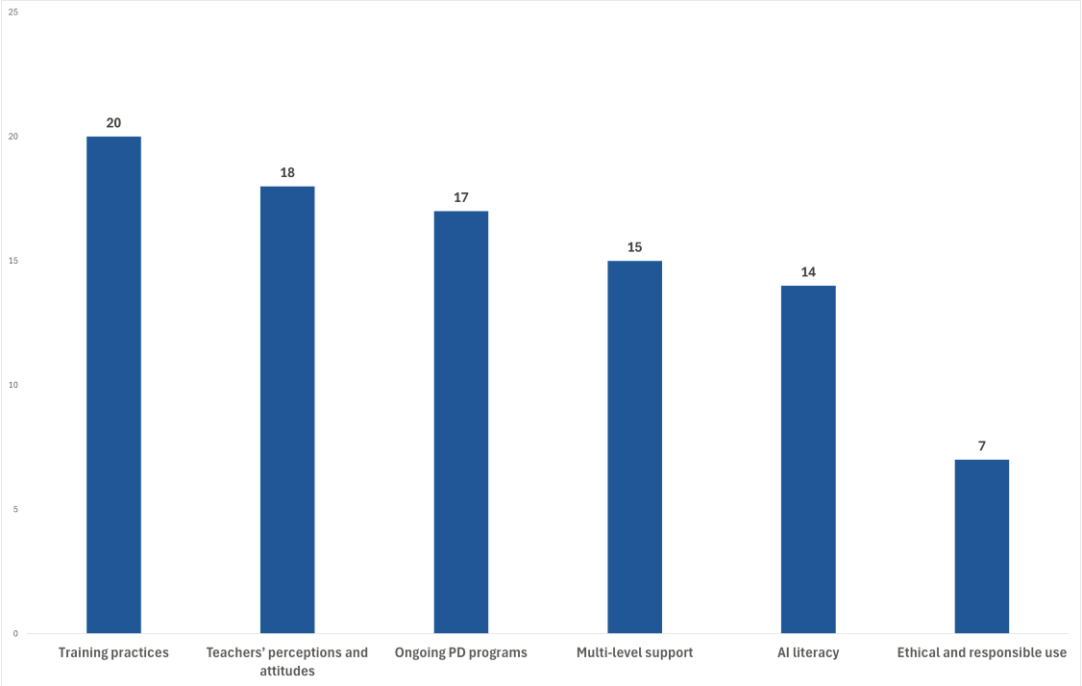


Figure 2. Teachers' professional development needs.

4.1. Training Practices

Table 3 presents the corresponding codes of this theme with the related references. Teacher training practices were the most mentioned theme in the empirical studies and consist of the ways in which teachers were trained for AI integration in school settings.

https://doi.org/10.3390/computers15010049

These practices were often constrained by existing gaps in teachers' AI literacy, which are analyzed in detail in Section 4.5.

Table 3. Codes in category: training practices.

4.1.1. Professional Learning Communities (n = 10)	[31–40]
4.1.2. Self-Reflection and Prompting (n = 4)	[41–44]
4.1.3. Case-Based and Application-Focused Learning (n = 3)	[45–47]
4.1.4. Differentiation in Level, Subject and Needs (n = 3)	[48–50]

Note: All these codes can be found in various empirical studies.

4.1.1. Professional Learning Communities

There seems to be a strong convergence among studies for the improvement of AI integration through teachers participating in collaborative practices, even though existing knowledge gaps about AI can limit the high motivation for sharing AI practices. Institutional and social norms influenced teachers and predicted AI readiness, with some studies highlighting structured CoP and others informal sharing networks.

Teachers displayed a keen interest in enhancing their proficiency in AI and facilitated the creation of networks to share good practice examples [38] and exchange knowledge and experiences regarding benefits, challenges, and solutions with other instructors [31]. To reduce resistance in integrating AI, communities of practice (CoP) [32,33] and peer collaboration [34,39] could help teachers share best practices and benefit from mutual learning. CoP and group activities with teachers who share common interests and concerns were identified as professional learning opportunities that could enhance understanding of AI, a common barrier, generate new levels of knowledge, and translate into new individual practices [40]. In addition, mentorship in learning communities and teacher-led resource-building networks that share AI use cases and guidebooks could support them in integrating AI into their teaching methods [37]. A collaborative approach to AI integration in schools, involving colleagues, administrators, and other educational stakeholders was indicated, as subjective norms significantly predicted AI readiness [35]. Additionally, opportunities to learn from others' feedback and PD experts could improve teachers' instructional practices and the quality of lecture design [39]. Teachers also stated that they wanted to be part of sharing knowledge about best practices, student education, expectations, ways to use ChatGPT for efficiency, lesson planning, student growth, verifying the academic integrity of papers, and other subjects that can benefit from ChatGPT [36]. The positive outcomes of such collaborative learning environments (CoP) in PD, where teachers and administrators could discuss difficulties, solve problems, and gain help in understanding and applying AI, were enhanced individual learning, collective growth, the promotion of solidarity, and effective instructional strategies [33].

4.1.2. Self-Reflection and Prompting

For the effectiveness of PD, there is a strong indication from the studies that GenAI is of great importance, as teachers should exercise their prompt engineering practices and construct their knowledge with AI-driven reflection, aligning with constructivist learning theories. On the contrary, despite being efficient, AI reflection depends on high-quality and contextualized data, making accuracy and pedagogical relevance challenging, while its use differs between experienced and novice teachers. Furthermore, it is obvious that knowledge about balancing autonomy and support regarding AI-enhanced PD is limited.

The importance of teachers' expertise in utilizing GenAI and designing effective prompts and activities with accuracy and efficiency was highlighted, as PD programs played a significant role in this regard [43]. Teachers used AI tools like ChatGPT for contin-

https://doi.org/10.3390/computers15010049

uous self-reflection, evaluating their teaching practices and enhancing professional growth. Teacher–AI collaboration (TAC) reflection was more flexible, less time-consuming, and self-directed compared with reflective practices that included class observation and consultation with colleagues [41]. However, more support is needed in this area, as this type of reflection requires high-quality data on practices, students, and outcomes, which must be analyzed thoroughly and contextualized for specific instructional settings. Prompting ChatGPT helped introspection, reflection, and an increased desire to evolve teaching practices, utilizing constructivist approaches for assimilating new information, and as participants clearly mentioned, through continuous cycles of questioning, reflecting, and adapting, they constructed knowledge [42]. This practice signified continuous adjustment of technological, pedagogical, and content knowledge, making connections to the TPACK framework and promoting holistic PD [42]. The benefits of chatting with conversational agents were also visible when using ChatPDF, especially for novice teachers, to read materials, gain pedagogical knowledge, receive practical guidance, and facilitate ongoing PD [44].

4.1.3. Case-Based and Application-Focused Learning

For the improvement of teachers' AI integration practices, case-based learning, hands-on activities, and real-world problem-solving are required to balance theoretical instruction with practical application. However, there are some inconsistencies among studies about the degree of structure needed in learning experiences to optimize teacher learning and efficacy.

Teachers who participated in PD training benefited most from real-world problem-solving and case-based learning, especially when these were balanced with direct instructions and lectures that helped them process knowledge about the technology and integrate it with prior pedagogical and content knowledge [46]. The researchers who ran the PD also recommended presenting teachers with additional cases and well-structured problems between lectures and case-based discussions, helping them apply new theories and tools in context and develop solutions [46]. Moderately structured and ill-structured problems could be used to engage teachers in solving, leading to more AI-related problem-finding and problem-solving discussions, while keeping a balance between direct instruction and case-based dialog in AI teacher PD programs [46]. To increase Computer Science teachers' AI teaching efficacy beliefs, TPACK-based PD programs could be employed with group-based lesson planning focused on project-based learning and hands-on activities, which could have a positive impact on teaching practices and students' learning outcomes [47]. Similarly, ref. [45] proposed that teacher PD activities could guide them to create self-directed learning tasks with AI, as this technology provides immediate feedback and suggestions to remediate unsuccessful learning.

4.1.4. Differentiation in Level, Subject, and Needs

Teachers require tailored PD rather than uniform approaches for efficient AI implementation in the classroom, with the focus on technical training to overcome technostress and limited understanding barriers. However, practical knowledge for utilizing AI-powered tools alone is not sufficient for pedagogically sound integration in school settings.

The need for tailored support to address challenges and PD opportunities for teachers to enhance their practices [49], as well as AI integration strategies in regard to educational stage, was underscored due to differences in AI implementation frequency in teaching across primary, secondary, and higher education [48]. Most teachers identified technical training as their primary need for effective ChatGPT implementation, with pedagogical training coming second in the instructional application of ChatGPT in mathematics education, underscoring that the greatest barrier was teachers' comfort with and understanding

https://doi.org/10.3390/computers15010049

of the AI tool, while simultaneously recognizing that operational knowledge of the tool was not enough [50]. The need for targeted exposure and training among mathematics teachers to raise awareness, deepen understanding, and enhance practical knowledge of AI tool application in classrooms was obvious [50].

4.2. Teachers' Perceptions and Attitudes

Table 4 presents this theme and the corresponding codes, as included in various empirical studies one or more times. It seems that negative perceptions of AI, such as teachers' fear of being replaced by it, anxiety, disbelief, and dishonesty, might be linked with misconceptions about AI's capabilities or previous negative experiences and play a significant role in challenging AI adoption practices. The fear of replacement by AI stems from concerns about loss of professional autonomy, uncertainty about the changing role of the teacher, and a limited understanding of how AI is intended to support rather than replace teaching. PD could shift these teachers' beliefs and emotions and enhance autonomy and self-efficacy, but the design of effective programs that address technical, affective, and pedagogical aspects can be difficult, yet essential for modern education.

Table 4. Codes in category: teachers' perceptions and attitudes.

AI knowledge AI understanding Autonomy Awareness Behavioral intentions to learn AI Beliefs and attitudes Conceptions of AI education Familiarity Self-efficacy Trust in AI Understanding of AI	(n = 11)	[31,32,34,39,46–48,51–54]
Anxiety Disbelief Fear of being replaced by AI Misconceptions	(n = 5)	[5,51,55–57]
Need for PD	(n = 5)	[5,34,35,58,59]

Note: All these codes can be found in various empirical studies.

Teachers demonstrated disbelief in AI, which may come from previous negative experiences and a lack of training [5]. Fear of replacement by AI was mentioned in many studies and persisted among teachers who agreed that ChatGPT cannot fulfill emotional bonding with students, highlighting their critical roles such as role modeling, individualized teaching, and empathy [56]. In ref. [55], participants were afraid of ChatGPT, which they had connected with cheating, and expressed feelings of anxiety, agreeing that it could cause more harm than good if left without PD and guidance. PD programs could help minimize or eliminate K-12 teachers' anxiety about the complications and unpredictability of AI and enhance their self-efficacy and behavioral intentions to learn about AI [51]. The exploration of ways to implement PD programs that teach K-12 teachers about AI remains a big challenge and complex task but is necessary to enhance knowledge and self-efficacy regarding, understanding of, and trust in AI, while addressing common misconceptions [57].

Training programs that were based on actionable frameworks that guide the integration of AI into teaching and learning without putting aside the human elements supported teachers' autonomy, enhanced self-efficacy in AI use, favored positive attitudes, had a significant effect on the quality of lectures, and improved adoption [39,46,47,52,54]. In-service

https://doi.org/10.3390/computers15010049

training for the practical application of AI through experiencing its use could be reflected in teachers' beliefs and attitudes, and therefore made it more likely for teachers to implement AI in their courses [48]. The findings pointed out that comprehensive and targeted PD programs should not only enhance TPACK but also address TAM (Technology Acceptance Model) factors such as teachers' attitudes to bridge the gap between knowledge and integration of AI, as positive attitudes towards it were very crucial for adoption [53]. Training courses for familiarization with tools and procedures, utilization techniques, and effective strategies for ChatGPT demonstrated that they could ensure awareness among female mathematics teachers in Saudi Arabia in a sex-oriented study [31]. In some cases, awareness and familiarity with AI-based tools came from various channels, including personal interest, workshops, and colleagues, as exposure to educational technology and training was insufficient [34].

The conceptions of AI education, highlighted by teachers' internal barriers, were very important in informing teachers' design and integration of AI in their classrooms [32]. Therefore, it was necessary to debunk myths about AI being a threat to humans and underscore its potential by explaining to teachers the technology behind its decision-making capabilities and gaining their trust to use it with their students [5]. In most cases, teachers expressed the need for PD and interventions to improve their AI integration [35,58], stated that a modern educator should master AI technologies to use them effectively [5], shared strategies to overcome challenges in problem-solving capabilities in mathematics education [34], and, even though some were experts, expressed the need to learn more, update their knowledge, and be advised by experts in multiple fields and subjects [59]. These perceptions are closely linked to the availability and quality of PD opportunities, which are discussed in detail in Section 4.3. The significance of these patterns is undeniable, as recognizing them allows us to explain why certain perceptions and negative feelings about AI exist among teachers and to seek ways to change them into more positive views and attitudes.

4.3. Ongoing PD Programs

Table 5 presents this theme and the corresponding codes as they were found in the empirical studies one or more times. In this theme, there is also repeated evidence that structured and continuous PD can facilitate teachers' AI integration by enhancing conceptual understanding, strategies, and the capacity for AI utilization across educational levels and subjects. The emerging divergences, however, concern the duration and design of PD, and inconsistencies between the acknowledged necessity of PD and the institutional support for implementation in practice.

The importance of providing professional training about GenAI to in-service teachers was highlighted, as study results showed increased conceptual understanding, improved strategies, and enhanced ability among teachers to integrate AI into lesson planning [52]. Continuous PD [32] and mentoring services for teachers could increase their competences in pedagogically integrating AI tools, such as ChatGPT, under Sustainable Development Goal 4.c.1 (SDG 4.c.1) [56]. Structured, continuous, context-centric, and on-demand PD programs for all levels of education [62], ensured effective AI implementation, especially if they were aligned with the needs of teachers, including their subject areas, curriculum objectives, and digital skills [5]. Results indicated an urgent need for structured PD to bridge the gap in familiarity with AI in mathematics instruction and to provide skills and confidence for pedagogical integration, addressing skepticism about ChatGPT in math education and education more broadly [36,50]. It was underlined by ref. [50] that enhancement of learning experiences must be the target of PD initiatives that will not only increase teachers' knowledge and skills for the integration of AI tools such as ChatGPT but will also demystify

https://doi.org/10.3390/computers15010049

AI and encourage critical engagement with technology, providing practical strategies to enrich mathematics education. In ref. [33], the importance of structured training and information curation in PD programs, as well as the critical evaluation of AI sources, was also mentioned, empowering teachers to rely on themselves when assessing knowledge and on clear guidelines for AI applications.

Table 5. Codes in category: ongoing PD programs.

Benefits (n = 17)	Address skepticism AI literacy Conceptual understanding Confidence Demystify AI Effective AI implementation Efficient assessment Encourage critical engagement Enhanced ability to integrate AI Enhanced knowledge Enhanced learning experiences Familiarity with AI Improved strategies Increased competences Increased self-efficacy Increased skills Maximum efficacy	[5,46,52,56,60–62]
Type (n = 9)	Additional PD Context-centric PD Continuous PD Long-term PD On-demand PD Ongoing training Short-term PD Structured PD Targeted PD	[32,33,38,46,50,55,57,62,63]
Challenges (n = 4)	Constant engagement Critical evaluation of AI sources Information curation Lack of policies	[33,41,55]
Methods (n = 4)	Case-based learning Implementation of AI in curriculum Learning community Practical application	[46,63,64]
Negative aspects (n = 3)	Policy-related courses Professors' lack of knowledge Theory-driven courses	[64]

Note: All these codes can be found in various empirical studies.

Ref. [46] explained the structure of a PD program about AI, which should have four domains to increase self-efficacy of participants: Knowledge and Understanding of AI (foundational knowledge of AI concepts and technologies), Applying AI Applications (employing AI in instructional activities), Evaluation of AI Applications (implications, ethical considerations, guidelines), and AI Ethics (data privacy, plagiarism, ethical guidelines, responsible use). They also suggested short-term introductory PD programs with case-based learning to introduce AI basics and its advantages and disadvantages, to provoke teachers' desire to increase their AI literacy, followed by extended AI teacher PD programs according

https://doi.org/10.3390/computers15010049

to their technical skills, subjects, and experience to support AI integration [46]. On the other hand, teachers saw short-term programs as insufficient for the learning process and emphasized the benefits of long-term PD, which included enhanced knowledge and skills in AI, space for a learning community in schools, and practical classroom application, while also mentioning negative aspects of teacher training programs such as impractical, theory-driven courses, redundant policy-related courses, and professors' lack of understanding of school realities [64].

Teachers recognized that ongoing training [38] was necessary for AI usage optimization and maximum efficacy [60], agreeing that additional PD, regardless of mixed experiences with district policies or the lack thereof across participants, would be useful [55]. Targeted PD programs could also highlight the long-term benefits of AI in education, in contexts where teachers emphasize virtues focused on the future [57]. Furthermore, Teacher–AI collaboration (TAC) required constant engagement for teachers to keep up with the latest developments in AI use, understanding ways to combine content, pedagogy, and technology [41], while training institutions also had to implement AI technologies more extensively into their curricula, along with application training [63]. Assessment processes could also become more efficient with AI, and PD was crucial in motivating teachers to accomplish this, primarily by developing AI knowledge and understanding and ensuring that teaching changes apply to all pedagogical possibilities [61]. Ongoing PD programs are very important for teachers' efficiency, and their form and methodology must consider the context to which they refer to be successful. In addition, several studies reported that ongoing PD played a mediating role in addressing previously identified anxieties and misconceptions toward AI (see Section 4.2).

4.4. Multi-Level Support

Table 6 presents this theme and the corresponding codes, as they were included in various empirical studies one or more times. Teacher readiness for AI integration cannot be sufficient without the existence of frameworks and policies to guide it, considering the lack of leadership and infrastructure support to confidently drive curriculum-aligned AI adoption. A culture of collaboration in organizations, and leadership vision and roles, can encourage innovation and reduce barriers, although material support and equitable investments are also needed.

Table 6. Codes in category: multi-level support.

Policies and guidelines (n = 8)	[32,36–38,43,50,55,65]
Leadership and organization (n = 8)	[32,45,53,59,66,67]
Support with PD (n = 6)	[38,40,59,65,67,68]
Technical support (n = 4)	[37,38,45,53]
Community (n = 1)	[53]

Note: All these codes can be found in various empirical studies.

Analytically, governmental support based on the Teacher Digital Competence framework (TDC) for new teaching methodologies was necessary, along with a human-centered use of AI to enhance teacher competencies [37]. External barriers made teachers consider a lack of AI guidance, as participants demonstrated mixed experiences regarding AI implementation, a concern that was addressed differently depending on schools and districts, with leaders creating a policy for or against the use of AI [55]. Clear guidelines and boundaries for using GenAI in educational contexts [43,65] and updated educational policies [38], following the needs of teachers, were in great demand to support sustainable AI integration in education [37]. Establishing clear curriculum guidelines could help teachers embrace AI

https://doi.org/10.3390/computers15010049

with confidence and develop enhanced learning environments [32], as many teachers had vague ideas for AI implementation and needed clarification [50]. Many teachers mentioned the desire for a district or school policy; therefore, government policymakers had to design the required guidelines [36].

Digital leadership and pedagogical and administrative support [53] could significantly contribute to learning about AI, enhancing teachers' competence and comfort in using AI-powered content, and aiding them in navigating difficulties with AI effectively, in an age-appropriate and engaging way [32]. The personal resources of leaders [69], and their vision and strategic direction to support innovation in school, were essential for AI integration and educational progress [66]. Therefore, school leaders could remove barriers and difficulties by creating supportive partnerships that helped and motivated teachers to share effective practices [32]. Teachers of the same subject could work as teaching groups and design common AI activities, raising their willingness to explore and incorporate AI, reducing stress, and fostering a pedagogical culture of collaboration, support, and innovation [67]. Also, on the other hand, teachers of different subjects could form professional learning communities to share knowledge and information and to recognize effective AI practices from a broader disciplinary perspective [59].

Teachers expressed the need for support in their own PD [65], mentioning insufficient PD in AI and skills that they would like to develop in order to work efficiently in technology-rich environments [68]. Many teachers were aware of the dynamic nature of AI and the need for refining their assessment strategies, but they also urged for PD [38] and training programs to continuously support them in developing AI knowledge and applying it in practice [59]. School support was very crucial for teachers to use AI technologies and develop within a suitably supported environment with appropriate PD activities [67]. It was highlighted in ref. [40] that the digital transformation of schools needed investments such as initiative piloting of AI and PD programs to enhance teachers' understanding of AI, as this was a common barrier connected with many issues. This level of support reinforces the need for structured and continuous PD programs (see Section 4.3).

However, support cannot be only organizational, in the form of learning communities or effective leadership, but must also come from available resources and infrastructure that can sustain the ethical and safe use of high-quality AI services [37]. In the study by ref. [45], it was stated that TPACK and UTAUT (Unified Theory of Acceptance and Use of Technology) could be synthesized in a new way, as the importance of facilitating conditions, like technical and expert support, was revealed in the early stages of AI adoption. The demand for robust and adequate technological infrastructure [38], ongoing technical support, and the prioritization of investments, especially in areas of low socio-economic status (SES), was underscored to ensure equity in access to AI technologies [53]. Additionally, schools could foster strong community ties and parental involvement in AI-related activities and discussions [53].

4.5. AI Literacy

Table 7 presents this theme and the corresponding codes, as they were included in various empirical studies one or more times. The primary challenge for AI integration, as many studies agree, is the substantial gap in AI literacy, which influences pedagogical application, confidence, self-efficacy, and willingness, even though it is not sufficient without instructional transformation. Traditional pedagogical frameworks like TPACK are frequently used, although they have limitations, which fuels the debate over which framework is the most appropriate for improving AI classroom implementation.

Elementary and secondary school teachers from various contexts and disciplines agreed that a big gap exists in understanding AI, identifying a strong demand for AI

https://doi.org/10.3390/computers15010049

literacy [72]. Adequate knowledge and literacy in utilizing ChatGPT for teachers and learners was underscored in ref. [56], while ref. [45] found that EFL teachers were more likely to continue integrating AI in their teaching when they had more technological knowledge about AI-based language applications. Addressing misconceptions about AI, equipping teachers with relevant knowledge, and aiming to improve basic understanding rather than providing unnecessary in-depth courses about AI, could enhance teachers' self-efficacy and help them understand it realistically as a tool with both benefits and disadvantages, making them more willing to try teaching with it [57].

Table 7. Codes in category: AI literacy.

AI knowledge (n = 10)	[45–47,53,54,56,57,67,70,71]
TPACK (n = 5)	[37,47,53,54,67]
AI understanding (n = 3)	[46,57,72]
AI literacy in PD (n = 2)	[5,51]
AI pedagogical implementation (n = 2)	[70,71]

Note: All these codes can be found in various empirical studies.

The integration of AI into existing pedagogical models and frameworks was mentioned in 43 articles with the following frequencies and percentages: TPACK (n = 11, 26%), TAM (n = 4, 9%), and TDC (Teacher Digital Competence) (n = 1, 2%). The study by ref. [37] highlighted the limitations of TPACK and suggested adopting TDC. Technological content knowledge (TCK) and technological pedagogical knowledge (TPK) had a strong impact on TPACK, emphasizing the need for comprehensive programs that could improve teachers' integrated knowledge, a crucial factor for AI implementation [53]. PD programs could be designed to promote teachers' AI-related TPACK and also provide them with the required technological knowledge so they could use technology's affordances for subject teaching and student learning through appropriate pedagogy, thereby improving AI adoption and integration in practice [67]. In ref. [47], a TPACK-based PD program improved Computer Science teachers' pedagogical content knowledge related to AI, their programming skills, an important element of technological AI-related knowledge, and teaching self-efficacy in terms of outcomes and beliefs (see Section 4.2). Likewise, another TPACK-based PD program had a positive impact on English language teachers' knowledge of AI-powered tools and on their self-efficacy in teaching with these tools [54].

The training of teachers on AI integration could not only focus on how the provided tools work but also on concepts like understanding AI functionalities, machine learning, and classification, as after a case-based AI PD program that aimed at improving AI literacy, teachers reported increased self-efficacy [46] The significance of AI literacy in the professional learning of K-12 teachers and the need to empower it as a central element were also stated in the study by ref. [51], while participants in the study by ref. [5] agreed that teacher training and PD courses should include modules that improve AI literacy, and that teachers should be trained in AI and provided with methods to apply it in the classroom. AI literacy would help teachers acquire foundational knowledge about AI, which would further affect its pedagogical application [70]. In ref. [71], it was emphasized that preparing teachers with AI knowledge as the basis of AI literacy was very important for understanding how they could transform aspects of that knowledge into pedagogical design. Teachers who had support to implement AI-related content into their pedagogical design were more effective in delivering knowledge to students, deepening their understanding of AI, and simultaneously enhancing their confidence in class management and in designing and teaching AI lessons [71]. AI literacy gaps underscore the importance of ongoing, structured PD programs (see Section 4.3).

https://doi.org/10.3390/computers15010049

4.6. Ethical and Responsible Use

Table 8 presents this theme and the corresponding codes, as they were included in various empirical studies one or more times. The ethical and safe use of AI is absolutely necessary; therefore, AI ethics should be integrated into teacher training programs as a cross-sectional principle. Ethics is a foundational component of AI competence, and the transformation of ethical principles into classroom practices is required. The safe and ethical AI use, focusing on transparency, responsibility, and equity to avoid potential risks from irresponsible use, can be promoted through policy guidelines, the creation of required infrastructure, and the implementation of AI ethics in teacher training [37,73]. In the study by ref. [71], it was suggested that PD programs should contain different types of knowledge and focus on empowering teachers to transform AI knowledge into pedagogical design and AI ethics education, which in turn can strengthen teachers' confidence in teaching AI and encourage their continued intention to learn and design AI courses. Integration of AI ethics in teacher training could improve their confidence and prepare them adequately [35], while also providing space to discuss social, ethical, and rights implications of AI and equipping teachers with the necessary knowledge [5]. Teachers need to be prepared to teach the ethical use of AI to students by integrating it into classroom activities and policies. As GenAI could play a central role in redefining educational practices, the development of comprehensive guidelines and ethical considerations to address this new reality with equity and alignment with educational goals is critical [43]. PD programs could also help teachers understand AI for social good and use it to help their students, while following AI ethical principles like justice and sustainability, as a part of their core content [51].

Table 8. Codes in category: ethical and responsible use.

Ensure ethical use of AI (n = 1)	[73]
AI ethical education (n = 1)	[71]
AI ethics and ethical principles AI for social good (n = 1)	[51]
AI ethics in PD (n = 1)	[35]
Strict regulations and expert controls for safe usage (n = 1)	[5]
Guidelines for safe, ethical, and equitable AI use Professional and ethical competencies in training (n = 1)	[37]
Guidelines to safeguard the educational process (n = 1)	[43]

Note: All these codes can be found in various empirical studies.

The application of AI is fraught with various interconnected ethical challenges, which can be described in terms of four main dimensions. The first is derived from privacy and data protection concerns arising from the extensive use of data, necessitating effective data governance practices. The second is found in bias and fairness concerns, pointing to the possibility of perpetuating and/or exacerbating existing inequalities in society through the use of unfair datasets or the development of unfair algorithms. Third, the need for transparency and explainability in the functioning of these AI systems is essential both for control and user trust. The fourth concerns accountability and responsibility, pinpointing efforts to identify individuals responsible for decisions made using AI. Such ethical considerations are ratified within larger global governance and policy regulatory frameworks. Specifically, the United Nations Educational Scientific and Cultural Organization (UNESCO) Recommendation on the Ethics of Artificial Intelligence [74] highlights irreducible principles of human rights protection, fairness, transparency, and accountability. These considerations and priorities are also encapsulated within other regulatory frameworks of global and regional promulgations, such as the Organization for Economic Cooperation

https://doi.org/10.3390/computers15010049

and Development (OECD) AI principles [75] and the EU regulation on artificial intelligence, the AI ACT [76], with an emphasis on effectively embedding ethical safeguards across the entire life cycle of AI systems, including design and development, use, and continuous assessment.

5. Discussion

In this systematic review, findings from empirical studies regarding teachers' training needs for future PD were categorized into themes. Across all themes, there were observable patterns like the requirement for professional training, knowledge gaps in AI literacy, and affective issues like anxiety and skepticism. To overcome redundancy, these points are highlighted within their pertinent sections. The heavy concentration of included studies on Asian countries has vital implications for the generalizability of the findings. The findings are of paramount significance for achieving insight into both the implementation and impact of AI on educational institutions, but the results are of limited universality in geographical terms, possibly representing institutional, cultural, or pedagogical specifics that are difficult to compare with others. The importance of differentiating applied educational contexts in analytical structures, evaluation schemes, digital maturity levels, and the regulation of digitalization also draws attention to the need for detailed interpretation of the findings, and possibly for expanding the results by involving a larger number of geographical regions in forthcoming studies concerning the use of AI in the mentioned establishments. The most frequently cited challenges to AI integration were related to teacher training [38,50,60] and AI literacy [46,49,72], therefore focusing on both will help overcome many practical obstacles. ChatGPT and GenAI in general accounted for a large percentage of the studies in our review, which is relevant with the fact that many existing systematic reviews focus on this specific AI tool. Thus, our findings are consistent with the ones in ref. [28] for improvements in GenAI for teaching and with the SWOT analysis of ref. [16] for ChatGPT's use in K-12 education. Many of the presented themes are linked to traditional technology integration frameworks such as TPACK, TDC, and SAMR; nevertheless, the findings suggest a paradigm shift towards human–AI collaboration in school settings, where AI enhances rather than replaces teachers' agency, creativity, and critical thinking for the benefit of their students, as previously seen in the proposed theoretical framework for human-centered AI (HCAI) in the systematic review by ref. [20].

This review highlights the importance of contextual factors for AI implementation and moves beyond traditional frameworks, agreeing with ref. [23], who noted differences in AI literacy definitions ranging from technical to socio-ethical frameworks, although a more balanced approach to their development would be better, considering not only educational level but also ethical aspects. Regarding ethical concerns, although they were the least cited category in our review, ref. [22] emphasized the critical construct of AI ethics in AI literacy and the need to develop an inclusive AI literacy competency framework. It is worth noting that issues of AI ethics do not seem to have been fully developed within an educational framework that would guide the integration of AI technology into learning, as the results of this review demonstrate that ethics was supported by the fewest references from empirical studies. This point is supported by the highlights of ref. [2], who emphasized that ethical and developmentally effective AI use requires teacher engagement as a mediating factor, hence reaffirming the importance of teacher PD and emphasizing pedagogy and ethics over automation. Violations of ethics, data security, privacy, academic integrity, and algorithmic bias must be addressed as a set of beliefs and values, even at a philosophical level [37,55,60,65]. Concerns about using AI and the role of teachers should be directed with clear guidelines and a critical approach to pedagogical and ethical issues such as

https://doi.org/10.3390/computers15010049

security, overdependence, bias and fairness, transparency and explainability, accountability, and privacy protection.

This study showed that AI literacy is impacted by many factors, reinforcing the findings of ref. [24], who mentioned, among others, the digital divide, perceptions and attitudes, and educational opportunities, highlighting that AI literacy is absent in teacher education, is crucial for AI learning, and should be the focus of PD. Different types of AI literacy should be identified and targeted through PD programs, such as foundational knowledge of AI terms and basic uses, and pedagogical literacy for teaching AI and using AI in teaching. In teacher training practices, self-reflection with AI was an effective and critical tool for in-service teachers' PD, which is relevant to the systematic review by ref. [19] on human–machine dialogic learning with teaching aids. In line with humanizing approaches to AI use in education, empirical evidence from primary school settings indicates that AI implementation remains teacher-led, as educators determine how the technology is implemented to education-related goals [2]. The need for ongoing PD programs and collaborative learning methods was eminent through most of our studies, agreeing with ref. [18], who underscored both the gap in teachers' PD and the potential of AI to transform educational systems. The findings of this review coincide with the ones from [26] and can be seen as complementary, adding insights from the theoretical foundations of AI integration that can be connected to empirical data for the proposed training framework. Ref. [25] stated the need for theoretical policy and teacher-driven initiatives for AI competence frameworks, based on empirical data from teachers and extending traditional technology integration theories. The practical use of the proposed framework is to serve teachers' educational goals with AI by providing guidance for its application not only in teacher training but also by bridging real classroom practices with AI technology integration. Teachers need to respond to the requirement to learn AI-based smart technologies, innovative ways of teaching and assessment, and the demand for classroom restructuring, as old pedagogies have become obsolete [70] and innovative education models should be developed [38].

5.1. Proposed Teacher Training Framework in AI Integration

Regarding this systematic review, we propose a potential framework for teachers' PD in AI integration. We explicitly cross-referenced each element of the framework with the corresponding research findings in the Section 4. The proposed framework has many practical implications for the design of AI-related teacher PD, extending from a purely technological view to a more pedagogically sound and ethical integration of AI. It is feasible to implement the proposed framework on a gradual scale based on levels of teacher readiness. From school leadership and policy-making standpoints, the proposed framework has serious implications for the role of leadership, infrastructure, and policies in the ethical integration of AI. From a research standpoint, the proposed framework also has many practical implications for the design of future PD. Analytically, this framework consists of four interrelated levels: (1) conditions for AI professional learning, (2) pedagogical design of AI-focused PD, (3) pedagogical AI integration in K-12 classrooms, and (4) ethical and sustainable embedding of AI.

5.1.1. Conditions for AI Professional Learning

The first level concerns the necessary conditions for teachers to participate effectively in PD programs. The findings of the review show that the absence of these conditions is a key reason for the failure of professional development interventions. At the primary level, preconditions such as basic AI knowledge and understanding, beliefs, self-efficacy, and support for teacher PD are required. AI literacy in this context [51] can incorporate AI knowledge [54] and AI understanding [37] to enable pedagogical AI implementation [70],

https://doi.org/10.3390/computers15010049

based on the traditional TPACK framework [67] and ethical AI use [71]. Additionally, focusing on teachers' perceptions and attitudes can give stakeholders a headstart in designing training, as these factors influence barriers and shortcomings [48,53]. Teachers' attitudes and perceptions can be recorded, including their fears, reservations, and expectations, emphasizing the importance of psychological/emotional factors and the need to conduct action research to adapt the training framework. Support for teachers' PD in AI, as already mentioned, must start at the individual level, but its multidimensional nature must also assign responsibilities where they belong. Difficulties such as lack of support [53] and the absence of guidelines [65], have challenged AI use in schools, therefore, teachers cannot be effective without the necessary infrastructure, technical support, or adequate forms of leadership at either the school or regional level [37,59,66]. The creation of a culture of cooperation, mutual support, and collective responsibility among teachers working in the same school is essential to transform the organization into a learning community. Additionally, networking between schools, universities, and other institutions could help, as national and international programs can provide the continuous feedback required, while education leaders are responsible for providing tools and guidelines, considering the important role of technical and organizational support.

5.1.2. Pedagogical Design of AI-Focused PD

The second level focuses on how PD programs for AI are designed. The results show that effective training is not based on the transmission of technical knowledge but on pedagogically documented learning experiences. Training practices, which were the most cited theme in this review, provided instances of professional learning communities [33], employment of self-reflection practices and prompting [43], and learning with differentiated approaches [50], in case-based and application-focused scenarios [46]. Using AI in PD practices could also be beneficial, especially when training teachers on how to use it, as AI-mediated reflection can adequately support them [42]. PD programs that incorporate reflection activities and hands-on practice to improve teachers' confidence and teaching efficiency could be designed, including courses about learning analytics. Professional learning communities (PLCs) can be created to aid teachers in sharing good practices, support each other with the challenges they are facing, and promoting innovative uses of AI, leading to increased implementation. Communities of practice, peer learning, and the development of micro-teaching can also make a decisive contribution to AI literacy. Apart from the methods already used for teacher training in the studies [33,42,46,50], experiential learning, problem-solving learning, and AI pedagogy with embedded ethics, are also recommended. Training methods should include collaborative exchange of knowledge and practices, tools that help with self-assessment, pilot interventions in the classroom, analysis of real situations and scenarios, experiential workshops with AI, innovative scenarios and simulations, and the cultivation of a critical approach (pedagogy and ethics of AI).

5.1.3. Pedagogical AI Integration in K-12 Classrooms

The third level concerns the transformation of teaching practices and is the point at which the effectiveness of training is essentially evaluated. Evaluation practices should not only concern classroom teachers but also all those who, due to their position, may hinder the integration of AI. Here, AI is integrated not as a substitute for the teacher but as a collaborative tool, which is why there needs to be a greater focus on GenAI, as this rapidly evolving technology offers many practical uses in content creation, lesson planning, curriculum development, assessment, feedback, and differentiation [43]. The employment of appropriate prompts when using conversational agents, which should be adequate and clear to have the expected results, is therefore crucial for teachers, who need practical

https://doi.org/10.3390/computers15010049

training in this area [49,55]. Some suggestions include developing prompting skills such as writing clear instructions, providing AI with information on the context, specifying the expected output format, tone, and style of writing, and the size/depth of analysis of the answer. Teachers can also practice with examples of good and bad prompts and apply them to their actual teaching needs, promoting 21st century skills and enhancing creativity, critical thinking, learner engagement, and motivation to participate [38,58,59,73]. Training must therefore include proposals for realistic, pedagogically approved practices that can be applied in real classroom conditions and in actual teaching contexts [70].

5.1.4. Ethical and Sustainable Embedding of AI

The fourth level concerns the long-term consolidation of AI in educational practice. The ethical use of AI (data, privacy, algorithmic biases, academic integrity) is not treated as an isolated issue but as a cross-cutting principle that permeates all levels of the framework. Although the focus of this review is mainly on in-service training for teachers, many suggestions can be used for teacher education in university and for graduate teachers. For teachers to understand the need for continuous PD, a distinction must be made between university education, pre-service training, and in-service training, and these must be linked to specific pedagogical objectives, durations, and forms. The structure of PD programs must be adapted to the context in terms of duration and format and should always align with the targeted teacher group [67,68,70]. Furthermore, for the effective integration of AI into teaching and learning, teachers need to attend training tailored to their specific needs [5,57]. Refs. [61,62] expressed that teachers benefited the most from ongoing PD programs, which led to effective AI implementation and improved skills. Continuous PD with a holistic approach that can transform teaching and learning processes should be the long-term goal of any intervention, rather than scattered training programs without specific pedagogical and didactic objectives. A barrage of tools and superficial practice is not effective; instead, long-term strategies and a holistic approach that analyzes educational needs, skills, attitudes, and perceptions to develop personalized learning paths. Training evaluation and feedback processes should take place at all levels: individual (teacher and learning outcomes) and organizational (school and regional levels), and everyone should participate, not just classroom teachers. PD strategies could be defined for every school community, and training should be integrated into daily routines rather than treated as an exceptional intervention.

5.2. Limitations and Future Recommendations

The research presented in this paper is subject to certain limitations. The diversity of study designs prevented the possibility of referring to a single standardized tool for assessing risk of bias. Future systematic reviews may include more fine-grained quality appraisals once the empirical literature on AI-related teacher PD continues to mature. Some of the categories may have conceptual overlaps, and the proposed framework does not suggest specific tools. In addition, some of our categories were strongly supported by many articles that highlighted general trends, while others had fewer references or specific cases, which nevertheless had to be mentioned. Future empirical studies could investigate these categories following qualitative approaches. In any case, some of the research data come from specific contexts and educational settings in certain areas with public or private school systems, policies, local curricula, and infrastructure that do not allow us to draw conclusions about all educational environments. In addition, most of the empirical studies come from the Asian continent. Therefore, future research is required to provide insights from other continents, countries, and educational systems. Similarly, researchers could explore how AI PD programs should be differentiated according to the

https://doi.org/10.3390/computers15010049

needs of teachers at different educational levels (primary, secondary, higher), subject areas (language, mathematics, arts, physics), different roles (teacher, consultant, leader), and specific contexts (educational policies, local communities, stakeholders, specific difficulties). The main themes for PD that emerged from this review can be used to design a research pathway on teachers' training needs in a specific context; then, by combining the results, a PD program could be organized, implemented, evaluated, and redesigned.

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

Funding: This research received no external funding.

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

Acknowledgments: This paper has been supported by the funding programme "MEDICUS" of the University of Patras.

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

Appendix A

Table A1. Empirical studies.

No.	Reference	Country–Region	Research Aim
1	[31]	Saudi Arabia	Propose a conceptual framework intended to leverage the capabilities of the ChatGPT artificial intelligence model to enhance the creative teaching proficiencies of secondary school mathematics teachers.
2	[60]	Saudi Arabia	Investigate teachers' perspectives and experiences with AI-powered technologies for creating customized learning materials and resources for students.
3	[45]	China	Investigate EFL teachers' perceptions, knowledge, and behavioral intention to use AI to support teaching and learning of English in middle schools.
4	[73]	Iran	Uncover how Iranian EFL teachers' ChatGPT-driven collaborative reflective practice (CRP), both independently and collaboratively, can contribute to their professional development.
5	[55]	USA	Understand the experiences and perceptions of teachers in a K-12 setting regarding the use of ChatGPT as a pedagogical tool.
6	[70]	Nepal	Examine teachers' awareness of AI, particularly its implications for teaching and learning among school teachers in Nepal.
7	[61]	Europe, Australasia, America, and Others—exact countries not stated	Examine the views of teachers from a diverse range of teaching levels, experience levels, discipline areas, and regions on the impact of AI on teaching and assessment; the ways they believe teaching and assessment should change; and the key motivations for changing their practices.

https://doi.org/10.3390/computers15010049

Table A1. *Cont.*

No.	Reference	Country–Region	Research Aim
8	[65]	Turkey	Investigate how school principals and teachers perceived the use of ChatGPT in education and reveal their attitudes towards using AI-based tools to facilitate the teaching–learning experience.
9	[71]	China	Assess teachers' efficacy in understanding AI and teaching AI, with additional considerations of promoting ethical awareness and designing socially beneficial AI applications.
10	[32]	Hong Kong	Explore the interplay between proactive digital leadership and the internal and external barriers teachers face in implementing AI.
11	[68]	Estonia	Explore teachers' perceptions about cutting-edge technologies (in this case, AI) and contextualize the results in the scope of Fairness, Accountability, Transparency, and Ethics (FATE).
12	[33]	USA	Provide effective PD programs for elementary school teachers and administrators working with bilingual children, with the goal of improving their AI literacy and positively influencing their attitudes towards AI integration in bilingual/ESL education.
13	[46]	USA	Examine the influence of a case-based AI professional development (PD) program on AI integration strategies and AI literacy.
14	[58]	Ukraine	Determine how actively AI and AI-assisted possibilities are used in Ukrainian school education and assess the attitudes of both Ukrainian students and teachers toward the use of AI in teaching and learning.
15	[51]	China	Examine the power of AI literacy and explore the determinants of behavioral intentions to learn AI among K-12 teachers.
16	[34]	Nigeria, South Africa	Explore teachers' awareness, utilization, and perceptions of ChatGPT.
17	[35]	Kenya, East Africa	Assess Kenyan K-12 in-service teachers' confidence in AI, attitudes toward AI, AI ethics, subjective norms, perceived threats, and readiness to teach AI and assess how these factors influence their readiness to teach AI.
18	[48]	Cyprus	Level of awareness regarding the use of AI.
19	[66]	Indonesia	Explore the intricate interplay between principal leadership styles, teacher roles, and the implementation of AI-based 'Merdeka' curriculum initiatives within vocational high schools across Indonesia.
20	[36]	USA	Determine the current perceptions and uses of ChatGPT in K-12 settings.
21	[37]	Republic of Korea	Explore the educational uses, considerations, and changes in social and educational environments in the transformational era of GenAI.
22	[49]	Turkey	Examine two key aspects: (i) how teachers adapt curricula within a specific framework, focusing on traditional adaptation patterns; and (ii) how GenAI tools, especially ChatGPT, can improve curriculum adaptation by exploring teachers' experiences in using AI to tailor educational content.
23	[41]	USA	Explore different types of Teacher–AI Collaboration and the potential benefits and obstacles of TAC.

https://doi.org/10.3390/computers15010049

Table A1. *Cont.*

No.	Reference	Country–Region	Research Aim
24	[59]	Republic of Korea	Explore teachers' perspectives on (1) curriculum design, (2) student–AI interaction, and (3) learning environments required to design student–AI collaboration (SAC) in learning, and (4) how SAC would evolve.
25	[38]	Turkey	Identify teachers' views on their digital skills in research studies using AI tools.
26	[52]	China	Propose a human-centered learning and teaching framework (HCLTF).
27	[64]	Republic of Korea	Understand in-service teachers' perceptions regarding AI education for teaching in schools and their AI teacher training programs.
28	[53]	China	Contribute empirical evidence to the growing body of knowledge regarding the factors that influence the utilization of AI tools in the educational sector, focusing on primary mathematics education.
29	[50]	The Philippines	Examine teachers' engagement with ChatGPT.
30	[62]	The Philippines	Contribute valuable insights to the ongoing discourse on integrating AI in language teaching and provide practical recommendations for teachers, curriculum developers, and educational policymakers seeking to harness AI's potential to enhance language learning experiences.
31	[39]	Vietnam	Investigate the relationship between teacher professional development, quality of lecture design, student engagement, teacher technical skills, pedagogical content knowledge, and teacher satisfaction in using artificial intelligence (AI)-powered facilitator for designing lectures.
32	[5]	The Russian Federation	Add to the growing volume of research that focuses on teachers' attitude towards AI, their views on its applicability in education, and the necessity to develop AI competences.
33	[47]	China	To promote CS teachers' AI teaching competency, a professional development (PD) program based on the Technological Pedagogical Content Knowledge (TPACK) framework was intentionally designed in this research, and its effectiveness was examined.
34	[42]	Vietnam	Delve into the experiences, perceptions, and outcomes associated with integrating ChatGPT into the TPD framework.
35	[72]	Taiwan	Explore how AI could enrich the physical education curriculum in elementary school.
36	[63]	Kazakhstan	Evaluate students' effectiveness in using AI.
37	[56]	Turkey	Evaluate the potential of using ChatGPT at the primary school level from the teachers' perspective within a sustainability framework.
38	[43]	South Africa	Explore how a specific group of teachers partner with GenAI tools, particularly ChatGPT, to complement and enhance their teaching.
39	[57]	Sweden, Norway, Israel, Japan, USA, and Brazil	Improve understanding of factors that influence teachers' adoption of AI-EdTech, investigating teachers' trust in AI-EdTech, and two of its antecedents, perceived benefits and concerns about AI-EdTech, in the context of K-12 education in six countries.

https://doi.org/10.3390/computers15010049

Table A1. *Cont.*

No.	Reference	Country–Region	Research Aim
40	[67]	China	Explore predictors of teachers' behavioral intentions to design AI-assisted learning and examine the structural relationships among these factors, by constructing a structural model of AI literacy, Technological Pedagogical Content Knowledge (TPACK), technostress, school support, teacher agency, teacher autonomy, and behavioral intentions.
41	[40]	China	Understand barriers involved in how AI is interpreted and the realities Hong Kong K-12 schools face, strengthen knowledge of current obstructions, and generate strategies that could help Hong Kong K-12 schools incorporate AI more effectively.
42	[44]	China	Explore how teachers in an English as a foreign language (EFL) context perceived and use a GenAI-based tool, ChatPDF, to develop materials for reading lessons
43	[54]	Indonesia	Assess the impact of the PD program on various aspects of AI teaching competence, including AI-powered tools knowledge test, teaching skills related to AI-powered tools, and AI-powered tools teaching self-efficacy.

Table A2. Studies per theme.

Training practices (n = 20)	Professional learning communities (n = 10)	[31–40]
	Self-reflection and prompting (n = 4)	[41–44]
	Case-based and application-focused learning (n = 3)	[45–47]
	Differentiation in level, subject, and needs (n = 3)	[48–50]
Teachers' perceptions and attitudes (n = 18)		[5,31,32,34,35,39,46–48,51–59]
Ongoing PD programs (n = 17)		[5,32,33,36,38,41,46,50,52,55–57,60–64]
Multi-level support (n = 15)		[32,36–38,40,43,45,50,53,55,59,65–68]
AI literacy (n = 14)		[5,37,45–47,51,53,54,56,57,67,70–72]
Ethical and responsible use (n = 7)		[5,35,37,43,51,71,73]

Note: All these codes can be found in various empirical studies.

Table A3. Previous systematic reviews and their main findings.

Authors	Main Findings
[15]	• Teachers can leverage AI chatbots to enhance their instruction, provide personalized support, and save time on routine tasks. • Concerns about reliability, accuracy, fairness, and ethical issues. • Offer customized feedback and instructional support, increasing student engagement and information retention, but not emotional support. • Educational institutions should implement preventative measures, including creating awareness among students and offering PD training for teachers.

https://doi.org/10.3390/computers15010049

Table A3. *Cont.*

Authors	Main Findings
[16]	Using ChatGPT in K-12 education (SWOT Analysis): • Strengths: personalized learning experiences, student achievement, personalized feedback, enhanced learning experiences, capability in various subjects, efficiency, time-saving attributes, ability to overcome language barriers, productivity, motivation and engagement, user-friendliness, conceptual understanding, language skills development, critical thinking, task correctness and adequacy, and high system usability. • Weaknesses: output quality, low task specificity, inability to handle certain questions, limited reasoning abilities, low productivity due to over-reliance, limited understanding, lack of contextual understanding, predictability, recognition of accents and dialects, slow response times, and lack of internet access for verification. • Opportunities: personalized learning applications, differentiated instruction, support for teachers to generate tasks and assessments, clear policies of use, critical thinking, revolutionizing teaching approaches, AI-supported curriculum development, best practice, and train teachers. • Threats: data privacy and ethical concerns, academic dishonesty, decreased productivity, prompt engineering, inaccurate information, lack of deep understanding, diminished interaction, superficial learning, harmful stereotypes, lack of teacher awareness, and users' potential disinterest.
[17]	• Ethical challenges from GenAI misinformation and disruption of student–teacher relationships, requiring accessible and inclusive programs to reduce inequalities. • Frameworks must consider legal, ethical, and technological aspects, with international regulations playing a key role. • GenAI offers opportunities to personalize and optimize teaching but poses challenges that require training programs for teachers.
[18]	• Need for ongoing PD for teachers to use new technologies with collaborative, innovative, inclusive, and transparent design. • There are ethical and privacy issues related to data security and student information, and a lack of practical examples that can inform implementation. • Challenges involve developing suitable curriculum and testing that incorporate AI tools in both an ethical and proper manner. • Development of appropriate and pedagogical tools for continuous review of teaching methodology and integration of technologies that enable diverse learning styles.
[19]	• Conceptual framework for Human–Machine Dialogic Learning (HMDL) in teacher education with three elements: Pedagogy (instructional strategies), Learning Environment (content and context), and Technological Platforms (pedagogical platform for human–machine interaction).
[20]	• The field of Human-Centered Artificial Intelligence (HCAI) is mainly characterized by theoretical studies, reflecting conceptual frameworks and model developments. • Dimensions covered in this literature include practical applications of HCAI, construction of theoretical models, integration of AI technology into educational systems, pedagogical approaches supported by HCAI, and ethical or value considerations. • The paper conceptualizes the HCAI evolution in education into a four-stage process, starting from genesis through convergence to place human needs at the center of mature integrated systems. • HCAI is applied in various educational scenarios such as personalized learning and teacher support. Ethical and value considerations attract attention with a view to ensure that technology aligns with human dignity, rights, and societal values.

https://doi.org/10.3390/computers15010049

Table A3. *Cont.*

Authors	Main Findings
[21]	• Studies on AI ethics in education explored themes such as Fairness and Equity, Privacy and Security, Non-maleficence and Beneficence, Agency and Autonomy, and Transparency and Intelligibility. • Fairness and equity are a central focus, with articles discussing personalized instructions, quality of education, fair learning opportunities, and social justice. • Privacy and security concerns governing data, including its collection, management, storage, access, and usage, involving the privacy of learners and teachers. • Multi-stakeholder collaborative efforts are needed to ensure that AI use does not dehumanize learning and has sustainable and positive effects on education.
[22]	• AI literacy involves a conceptual blend of digital literacy, computational thinking, critical data literacy, and AI ethics. • Utilization of constructivist methodologies results in positive academic, affective, and behavioral outcomes. • Inform teachers and policymakers about effective implementation while highlighting challenges like the need for a systematic AI curriculum and teacher PD.
[23]	• AI literacy definitions have varied from technical proficiency to socio-ethical frameworks; thus, a balanced approach is most obviously required. • AI literacy frameworks are technical or holistic, which need to be considered in curriculum development and integration across educational levels. • AI literacy is an important element in modern educational pedagogies for its ethical, social, and applicative dimensions.
[24]	• AI literacy involves knowing dimensions—understand, apply, and ethics—and stakeholder perceptions among teachers and parents. • AI literacy integration focuses on development—innovative teaching methods and influencing factors such as gender and grade levels. • AI literacy assessment: measurement (scale development and verification) and effects (behavioral intentions, learning efficacy). • Increased global attention towards AI in education, primarily at the K-12 level. • Factors that have an impact on AI literacy and stakeholder perceptions remain under-researched areas. • AI literacy development among K-12 teachers is scant yet an essential part of their AI education.
[25]	• AI Competence Frameworks for Teachers (AI CFTs) classified within Competence Construct Claims, providing a structured classification to understand their foci and characteristics. • The authors distinguished five types of frameworks: (1) existing competence model-oriented; (2) competence areas/domains-oriented; (3) process-driven (process competency model); (4) AI systems-driven (specific work/job model), and (5) competence level-driven (multidimensional). • The classification of AI Competence Frameworks (CFTs) can be instrumental for researchers and policymakers in comprehending approaches to teacher competence development.

https://doi.org/10.3390/computers15010049

Table A3. *Cont.*

Authors	Main Findings
[26]	Teachers' challenges with AI integration: • Technical expertise needs to develop into effective pedagogical practices. • Learning to navigate ethical considerations and contextual factors related to school culture and resources. • Struggle to understand and apply AI tools practically in instructional routines. • Lack of adequate awareness and guidance in addressing the ethical dilemmas related to AI use in the classroom. • Teachers' perceptions of AI capabilities create challenges of trust and effective implementation. • Technological proficiency and teaching contexts complicate the adoption process. • Organizational and infrastructural barriers including school culture, available resources, and administrative support. • As AI technologies continue to evolve, it will be increasingly important for teachers to play a key role in properly integrating such innovations into practice.

References

1. Aravantinos, S.; Lavidas, K.; Voulgari, I.; Papadakis, S.; Karalis, T.; Komis, V. Educational Approaches with AI in Primary School Settings: A Systematic Review of the Literature Available in Scopus. *Educ. Sci.* **2024**, *14*, 744. [CrossRef]
2. Voulgari, I.; Lavidas, K.; Aravantinos, S.; Sypsa, S.; Sfyroera, M. Exploring the Role, Implementation, and Educational Implications of AI in Early Childhood Education. In *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky*; Olney, A.M., Chounta, I.-A., Liu, Z., Santos, O.C., Bittencourt, I.I., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 29–37. [CrossRef]
3. Wijayanti, A.; Dwiningrum, S.I.A.; Saptono, B. Digital Literacy in Elementary Schools Post COVID-19: A Systematic Literature Review. *Int. J. Learn. Teach. Educ. Res.* **2024**, *23*, 96–112. [CrossRef]
4. Su, J.; Zhong, Y.; Ng, D.T.K. A meta-review of literature on educational approaches for teaching AI at the K-12 levels in the Asia-Pacific region. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100065. [CrossRef]
5. Sadykova, G.; Kayumova, A. Educators' Perception of Artificial Intelligence as Instructional Tool. *TEM J.* **2024**, *13*, 3194–3204. [CrossRef]
6. Miao, F.; Cukurova, M. *AI Competency Framework for Teachers*; United Nations Educational, Scientific and Cultural Organization: Paris, France, 2024. [CrossRef]
7. Ng, D.T.K.; Leung, J.K.L.; Chu, S.K.W.; Qiao, M.S. Conceptualizing AI literacy: An exploratory review. *Comput. Educ. Artif. Intell.* **2021**, *2*, 100041. [CrossRef]
8. Mishra, P.; Warr, M.; Islam, R. TPACK in the age of ChatGPT and Generative AI. *J. Digit. Learn. Teach. Educ.* **2023**, *39*, 235–251. [CrossRef]
9. Celik, I. Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Comput. Hum. Behav.* **2023**, *138*, 107468. [CrossRef]
10. Ning, Y.; Zhang, C.; Xu, B.; Zhou, Y.; Wijaya, T.T. Teachers' AI-TPACK: Exploring the Relationship between Knowledge Elements. *Sustainability* **2024**, *16*, 978. [CrossRef]
11. Zou, D.; Xie, H.; Kohnke, L. Navigating the Future: Establishing a Framework for Educators' Pedagogic Artificial Intelligence Competence. *Eur. J. Educ.* **2025**, *60*, e70117. [CrossRef]
12. Filiz, O.; Kaya, M.H.; Adiguzel, T. Teachers and AI: Understanding the factors influencing AI integration in K-12 education. *Educ. Inf. Technol.* **2025**, *30*, 17931–17967. [CrossRef]
13. Lin, C.-C.; Huang, A.Y.Q.; Lu, O.H.T. Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review. *Smart Learn. Environ.* **2023**, *10*, 41. [CrossRef]
14. Chiu, T.K.F.; Ahmad, Z.; Çoban, M. Development and validation of teacher artificial intelligence (AI) competence self-efficacy (TAICS) scale. *Educ. Inf. Technol.* **2025**, *30*, 6667–6685. [CrossRef]
15. Labadze, L.; Grigolia, M.; Machaidze, L. Role of AI chatbots in education: Systematic literature review. *Int. J. Educ. Technol. High. Educ.* **2023**, *20*, 56. [CrossRef]
16. Zhang, P.; Tur, G. A systematic review of ChatGPT use in K-12 education. *Eur. J. Educ.* **2024**, *59*, e12599. [CrossRef]
17. García-López, I.M.; Trujillo-Liñán, L. Ethical and regulatory challenges of Generative AI in education: A systematic review. *Front. Educ.* **2025**, *10*, 1565938. [CrossRef] [PubMed]

https://doi.org/10.3390/computers15010049

18. Marzano, D. Generative Artificial Intelligence (GAI) in Teaching and Learning Processes at the K-12 Level: A Systematic Review. *Technol. Knowl. Learn.* **2025**, *30*, 1–41. [CrossRef]
19. Wan, P.; Gu, X. Developing teachers' professional abilities: A systematic review of human-machine dialogic learning for teacher education. *Interact. Learn. Environ.* **2025**, *33*, 1–25. [CrossRef]
20. Chen, X.; Hu, Z.; Li, Y.; Wang, C. The journey of challenges and triumphs: A systematic literature review of the dynamic evolution of human-centered artificial intelligence in education. *Interact. Learn. Environ.* **2025**, *33*, 4906–4931. [CrossRef]
21. Fu, Y.; Weng, Z. Navigating the ethical terrain of AI in education: A systematic review on framing responsible human-centered AI practices. *Comput. Educ. Artif. Intell.* **2024**, *7*, 100306. [CrossRef]
22. Yim, I.H.Y.; Su, J. Artificial intelligence literacy education in primary schools: A review. *Int. J. Technol. Des. Educ.* **2025**, *35*, 2175–2204. [CrossRef]
23. Biagini, G. Towards an AI-Literate Future: A Systematic Literature Review Exploring Education, Ethics, and Applications. *Int. J. Artif. Intell. Educ.* **2025**, *35*, 2616–2666. [CrossRef]
24. Tan, Q.; Tang, X. Unveiling AI literacy in K-12 education: A systematic literature review of empirical research. *Interact. Learn. Environ.* **2025**, *33*, 5347–5363. [CrossRef]
25. Mikeladze, T.; Meijer, P.C.; Verhoeff, R.P. A comprehensive exploration of artificial intelligence competence frameworks for educators: A critical review. *Eur. J. Educ.* **2024**, *59*, e12663. [CrossRef]
26. Dogan, S.; Nalbantoglu, U.Y.; Celik, I.; Agacli Dogan, N. Artificial intelligence professional development: A systematic review of TPACK, designs, and effects for teacher learning. *Prof. Dev. Educ.* **2025**, *51*, 519–546. [CrossRef]
27. Martin, F.; Zhuang, M.; Schaefer, D. Systematic review of research on artificial intelligence in K-12 education (2017–2022). *Comput. Educ. Artif. Intell.* **2024**, *6*, 100195. [CrossRef]
28. Noroozi, O.; Soleimani, S.; Farrokhnia, M.; Banihashem, S.K. Generative AI in Education: Pedagogical, Theoretical, and Methodological Perspectives. *Int. J. Technol. Educ.* **2024**, *7*, 373–385. [CrossRef]
29. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
30. Braun, V.; Clarke, V. Thematic analysis. In *APA Handbook of Research Methods in Psychology*; Cooper, H., Camic, P.M., Long, D.L., Panter, A.T., Rindskopf, D., Sher, K.J., Eds.; Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological; American Psychological Association: Washington, DC, USA, 2012; Volume 2, pp. 57–71. [CrossRef]
31. Alhazzani, N.S. Enhancing mathematics teachers' pedagogical skills by using ChatGPT. *Int. J. Innov. Res. Sci. Stud.* **2024**, *7*, 1614–1626. [CrossRef]
32. Cheng, E.C.K.; Wang, T. Leading digital transformation and eliminating barriers for teachers to incorporate artificial intelligence in basic education in Hong Kong. *Comput. Educ. Artif. Intell.* **2023**, *5*, 100171. [CrossRef]
33. Davoodi, A. Crafting innovative paths in non-linear professional learning for bilingual education: The role of connectivism in the age of AI. *Prof. Dev. Educ.* **2024**, *51*, 434–450. [CrossRef]
34. Egara, F.O.; Mosimege, M. Exploring the Integration of Artificial Intelligence-Based ChatGPT into Mathematics Instruction: Perceptions, Challenges, and Implications for Educators. *Educ. Sci.* **2024**, *14*, 742. [CrossRef]
35. Fundi, M.; Sanusi, I.T.; Oyelere, S.S.; Ayere, M. Advancing AI education: Assessing Kenyan in-service teachers' preparedness for integrating artificial intelligence in competence-based curriculum. *Comput. Hum. Behav. Rep.* **2024**, *14*, 100412. [CrossRef]
36. Hays, L.; Jurkowski, O.; Sims, S.K. ChatGPT in K-12 Education. *TechTrends* **2024**, *68*, 281–294. [CrossRef]
37. Jang, H.; Choi, H. A Double-Edged Sword: Physics Educators' Perspectives on Utilizing ChatGPT and Its Future in Classrooms. *J. Sci. Educ. Technol.* **2025**, *34*, 267–283. [CrossRef]
38. Kokoç, M. Factors influencing K-12 teachers' experiences of using Generative AI Tools: Opportunities and barriers. *J. E-Learn. Knowl. Soc.* **2024**, *20*, 101–111. [CrossRef]
39. Nguyen, T.H.C. Exploring the Role of Artificial Intelligence-Powered Facilitator in Enhancing Digital Competencies of Primary School Teachers. *Eur. J. Educ. Res.* **2024**, *13*, 219–231. [CrossRef]
40. Wang, T.; Cheng, E.C.K. An investigation of barriers to Hong Kong K-12 schools incorporating Artificial Intelligence in education. *Comput. Educ. Artif. Intell.* **2021**, *2*, 100031. [CrossRef]
41. Kim, J. Types of teacher-AI collaboration in K-12 classroom instruction: Chinese teachers' perspective. *Educ. Inf. Technol.* **2024**, *29*, 17433–17465. [CrossRef]
42. Tran, T.T.; Pham, T.T.; Le, T.T. Transformative Interactions: ChatGPT's Role in Facilitating Professional Development Among Vietnamese English as a Foreign Language Teachers. *High. Learn. Res. Commun.* **2024**, *14*, 48–63. [CrossRef]
43. Van Den Berg, G. Generative AI and Educators: Partnering in Using Open Digital Content for Transforming Education. *Open Prax.* **2024**, *16*, 130–141. [CrossRef]
44. Xin, J.J. Investigating EFL teachers' use of generative AI to develop reading materials: A practice and perception study. *Lang. Teach. Res.* **2024**, *29*, 1–30. [CrossRef]

https://doi.org/10.3390/computers15010049

45. An, X.; Chai, C.S.; Li, Y.; Zhou, Y.; Shen, X.; Zheng, C.; Chen, M. Modeling English teachers' behavioral intention to use artificial intelligence in middle schools. *Educ. Inf. Technol.* **2023**, *28*, 5187–5208. [CrossRef]
46. Ding, A.-C.E.; Shi, L.; Yang, H.; Choi, I. Enhancing teacher AI literacy and integration through different types of cases in teacher professional development. *Comput. Educ. Open* **2024**, *6*, 100178. [CrossRef]
47. Sun, J.; Ma, H.; Zeng, Y.; Han, D.; Jin, Y. Promoting the AI teaching competency of K-12 computer science teachers: A TPACK-based professional development approach. *Educ. Inf. Technol.* **2023**, *28*, 1509–1533. [CrossRef]
48. Güneyli, A.; Burgul, N.S.; Dericioğlu, S.; Cenkova, N.; Becan, S.; Şimşek, Ş.E.; Güneralp, H. Exploring Teacher Awareness of Artificial Intelligence in Education: A Case Study from Northern Cyprus. *Eur. J. Investig. Health Psychol. Educ.* **2024**, *14*, 2358–2376. [CrossRef]
49. Karataş, F.; Eriçok, B.; Tanrikulu, L. Reshaping curriculum adaptation in the age of artificial intelligence: Mapping teachers' AI -driven curriculum adaptation patterns. *Br. Educ. Res. J.* **2025**, *51*, 154–180. [CrossRef]
50. Magat, R.J.B.; Sangalang, E.M. Teachers' Familiarity, Perceptions, and Training Needs on the Use of ChatGPT in Mathematics Instruction. *Int. J. Educ. Math. Sci. Technol.* **2024**, *12*, 1471–1487. [CrossRef]
51. Du, H.; Sun, Y.; Jiang, H.; Islam, A.Y.M.A.; Gu, X. Exploring the effects of AI literacy in teacher learning: An empirical study. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 559. [CrossRef]
52. Kong, S.-C.; Yang, Y. A Human-Centered Learning and Teaching Framework Using Generative Artificial Intelligence for Self-Regulated Learning Development Through Domain Knowledge Learning in K–12 Settings. *IEEE Trans. Learn. Technol.* **2024**, *17*, 1562–1573. [CrossRef]
53. Li, M. Integrating Artificial Intelligence in Primary Mathematics Education: Investigating Internal and External Influences on Teacher Adoption. *Int. J. Sci. Math. Educ.* **2025**, *23*, 1283–1308. [CrossRef]
54. Zulianti, H.; Hastuti, H.; Nurchurifiani, E.; Hastomo, T.; Maximilian, A.; Ajeng, G.D. Enhancing Novice EFL Teachers' Competency in AI-Powered Tools Through a TPACK-Based Professional Development Program. *World J. Engl. Lang.* **2024**, *15*, 117. [CrossRef]
55. Bateman, T. Teacher perspectives of ChatGPT as a pedagogical tool in the K-12 setting: A case study. *Qual. Assur. Educ.* **2025**, *33*, 203–217. [CrossRef]
56. Uğraş, H.; Uğraş, M.; Papadakis, S.; Kalogiannakis, M. ChatGPT-Supported Education in Primary Schools: The Potential of ChatGPT for Sustainable Practices. *Sustainability* **2024**, *16*, 9855. [CrossRef]
57. Viberg, O.; Cukurova, M.; Feldman-Maggor, Y.; Alexandron, G.; Shirai, S.; Kanemune, S.; Wasson, B.; Tømte, C.; Spikol, D.; Milrad, M.; et al. What Explains Teachers' Trust in AI in Education Across Six Countries? *Int. J. Artif. Intell. Educ.* **2025**, *35*, 1288–1316. [CrossRef]
58. Dovgyi, S.; Babiichuk, S.; Davybida, L.; Biletska, M. Teachers' and students' attitudes towards the use of artificial intelligence: All-ukrainian research. *Inf. Technol. Learn. Tools* **2024**, *104*, 197–215. [CrossRef]
59. Kim, J.; Lee, H.; Cho, Y.H. Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Educ. Inf. Technol.* **2022**, *27*, 6069–6104. [CrossRef]
60. Almuhanna, M.A. Teachers' perspectives of integrating AI-powered technologies in K-12 education for creating customized learning materials and resources. *Educ. Inf. Technol.* **2025**, *30*, 10343–10371. [CrossRef]
61. Bower, M.; Torrington, J.; Lai, J.W.M.; Petocz, P.; Alfano, M. How should we change teaching and assessment in response to increasingly powerful generative Artificial Intelligence? Outcomes of the ChatGPT teacher survey. *Educ. Inf. Technol.* **2024**, *29*, 15403–15439. [CrossRef]
62. Mananay, J.A. Integrating Artificial Intelligence (AI) in Language Teaching: Effectiveness, Challenges, and Strategies. *Int. J. Learn. Teach. Educ. Res.* **2024**, *23*, 361–382. [CrossRef]
63. Uaidullakyzy, E.; Oralbekova, A.K.; Dosbenbetova, A.S.; Yerubay, B.; Nauryzbayev, B.A.; Turmanov, R. Examination of students' success in the use of artificial intelligence. *S. Afr. J. Educ.* **2024**, *44*, 1–11. [CrossRef]
64. Lee, Y.-J.; Davis, R.O.; Ryu, J. Korean in-Service Teachers' Perceptions of Implementing Artificial Intelligence (AI) Education for Teaching in Schools and Their AI Teacher Training Programs. *Int. J. Inf. Educ. Technol.* **2024**, *14*, 214–219. [CrossRef]
65. Cetin, Y.; Taş, Ö.; Alakuş, H.; Kaplan, H.İ. Examining School Principals' and Teachers' Perceptions of Using ChatGPT in Education. *Educ. Process Int. J.* **2024**, *13*, 85–96. [CrossRef]
66. Halomoan, H.; Hakiki, M.; Ramadhan, M.A.; Hidayah, Y.; Fakhri, J.; Aljamaliah, S.N.M.; Abi Hamid, M. Integrating Principal Leadership and Teacher Roles with AI-Based "Merdeka" Curriculum Innovation: The Quantitative Research. *TEM J.* **2024**, *13*, 3397–3404. [CrossRef]
67. Wang, K.; Chai, C.-S.; Liang, J.-C.; Sang, G. Exploring teachers' behavioural intentions to design artificial intelligence-assisted learning in Chinese K–12 education. *Technol. Pedagog. Educ.* **2024**, *33*, 629–645. [CrossRef]
68. Chounta, I.-A.; Bardone, E.; Raudsep, A.; Pedaste, M. Exploring Teachers' Perceptions of Artificial Intelligence as a Tool to Support their Practice in Estonian K-12 Education. *Int. J. Artif. Intell. Educ.* **2022**, *32*, 725–755. [CrossRef]
69. Leithwood, K. The Personal Resources of Successful Leaders: A Narrative Review. *Educ. Sci.* **2023**, *13*, 932. [CrossRef]

https://doi.org/10.3390/computers15010049

70. Bohara, D.K.; Rana, K. Unmasking teachers' proficiency in harnessing Artificial Intelligence (AI) for transformative education. *SN Soc. Sci.* **2024**, *4*, 203. [CrossRef]
71. Chai, C.S.; Liang, S.; Wang, X. A survey study of chinese teachers' continuous intentions to teach artificial intelligence. *Educ. Inf. Technol.* **2024**, *29*, 14015–14034. [CrossRef]
72. Tseng, C.-W.; Chang, A.; Liu, E.-Y. *Detecting Fitted Strategies for Integrating AI Applications into Physical Education Using Fuzzy Logic (No. 10)*; ICIC International: Kumamoto, Japan, 2024. [CrossRef]
73. Arefian, M.H.; Çomoğlu, I.; Dikilitaş, K. Understanding EFL teachers' experiences of ChatGPT-driven collaborative reflective practice through a community of practice lens. *Innov. Lang. Learn. Teach.* **2024**, *19*, 1–16. [CrossRef]
74. United Nations Educational; Scientific and Cultural Organization. Recommendation on the Ethics of Artificial Intelligence. 2021. Available online: https://unesdoc.unesco.org/ark:/48223/pf0000381137 (accessed on 1 January 2026).
75. *OECD/LEGAL/0449*; Recommendation of the Council on Artificial Intelligence. Organization for Economic Co-operation and Development: Paris, France, 2024. Available online: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (accessed on 1 January 2026).
76. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Official Journal L 2024/1689. 2024. Available online: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (accessed on 1 January 2026).

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.

https://doi.org/10.3390/computers15010049

computers

Systematic Review

A Systematic Review of Mind Maps, STEM Education, Algorithmic and Procedural Learning

Chrysovalantis Kefalis [1,2], Constantine Skordoulis [2] and Athanasios Drigas [1,*]

[1] Net Media Lab & Mind & Brain R&D, Institute of Informatics & Telecommunications, National Centre of Scientific Research 'Demokritos', Patr. Gregoriou E & 27 Neapoleos Str., 15341 Agia Paraskevi, Greece; vkefalis@iit.demokritos.gr
[2] Department of Primary Education, National and Kapodistrian University of Athens, Marasli 4, 10676 Athens, Greece; kskordul@primedu.uoa.gr
* Correspondence: dr@iit.demokritos.gr

Abstract: This systematic review investigates the use of mind maps in STEM education, focusing on their application, effectiveness, and contextual factors. The main objectives were to examine whether mind maps are used as learning or assessment tools, the research designs employed, the type of interaction (individual vs. collaborative), and the format (digital vs. paper-based). Studies were identified through systematic searches in ERIC, Scopus, and Web of Science, including peer-reviewed journal articles published between 2019 and 2024. The inclusion criteria required empirical research studies using mind maps in STEM contexts with measurable outcomes related to learning or engagement. Studies without empirical data or not focused on STEM education were excluded. Fifty studies met the inclusion criteria. Most employed quasi-experimental designs (n = 29), including 22 with pre–post-test measurements. The mind maps were mainly used as learning tools (n = 40), in individual settings (n = 24), with student-generated (n = 36) and digital formats (n = 21) being most common. The reported outcomes included improved academic performance, conceptual understanding, critical thinking, and motivation and reduced cognitive load. The limitations included inconsistent reporting of the map types and theoretical underpinnings. The findings suggest that mind maps are effective tools for enhancing learning and engagement in STEM education and warrant broader pedagogical integration.

Keywords: mind maps; STEM education; concept mapping; learning outcomes; student engagement; cognitive skills; educational technology; algorithmic learning; procedural thinking; concept representation

1. Introduction

Algorithmic learning and procedural learning are essential components of STEM education, encompassing students' ability to understand and follow step-by-step processes, logical sequences, and structured problem-solving approaches. These skills are particularly emphasized in disciplines such as mathematics, computer science, and engineering [1,2]. Mind maps—by visually representing the relationships between concepts, steps, and sub-processes—offer a powerful way to scaffold these skills. Studies have shown that concept mapping supports the development of algorithmic thinking and procedural fluency by enhancing students' ability to organize sequential information and identify dependencies [3,4].

Mind maps are graphical tools designed to organize and structure information in a non-linear, interconnected manner. Introduced by Tony Buzan in the 1970s, mind maps

are often represented as tree-like structures, with a central idea branching out into related subtopics. Each branch may contain keywords, symbols, or images that reinforce learning and memory retention. According to Buzan (2006), mind maps utilize both verbal and visual elements to engage the whole brain, promoting comprehension and long-term recall [5]. These features have been confirmed in more recent empirical studies in STEM education [6]. Similar findings are reported in broader studies on digital learning tools that support cognitive and metacognitive development [7].

Mind maps play a significant role in STEM education by enhancing learning outcomes, fostering cognitive engagement, and improving one's conceptual understanding. The research indicates that mind maps help students structure their knowledge by visually organizing information, making connections between concepts, and developing problem-solving strategies [1,8]. One of the key benefits of mind maps in STEM education is their ability to improve students' knowledge retention and comprehension. Studies have shown that incorporating mind maps in mathematics and science courses leads to higher test scores and deeper conceptual understanding, particularly in areas requiring logical reasoning, such as geometry and problem-solving [6,9–11].

One of the key benefits of mind maps in STEM education is their ability to improve one's knowledge retention and comprehension. Studies have shown that incorporating mind maps in mathematics courses leads to higher test scores and a deeper conceptual understanding, particularly in areas requiring logical reasoning, such as geometry [6]. By visually representing relationships between abstract ideas, students can develop stronger analytical skills, which are crucial in problem-solving-based STEM disciplines [12].

From a cognitive perspective, mind maps support the activation of both hemispheres of the brain, engaging students in spatial reasoning and logical analyses simultaneously [6]. This balance is particularly useful in STEM subjects that demand a combination of sequential thinking (e.g., solving equations in mathematics) and holistic understanding (e.g., recognizing patterns in physics or engineering). Additionally, studies have linked concept mapping activities to improvements in metacognitive skills, allowing students to monitor and regulate their learning processes [12–14]. These benefits align with broader educational frameworks emphasizing the importance of digital tools in fostering metacognitive growth and emotional engagement in future-oriented learning environments [15].

Furthermore, mind maps enhance students' engagement and motivation by promoting active learning. In flipped classroom settings, for example, students who used concept maps to guide their learning activities reported higher levels of engagement and improved ability to link prior knowledge to new concepts [14,16,17]. Similarly, in engineering education, cognitive mapping techniques have been found to facilitate knowledge transfer and interdisciplinary thinking, which are essential for tackling complex STEM challenges [18].

Concept maps and mind maps are visual knowledge representation tools that are widely employed in STEM education. Although concept maps typically emphasize hierarchical or network relationships, and mind maps generally focus on radial structures for idea generation, these terms are sometimes used interchangeably or without clear differentiation in educational research [1,2,19,20]. Therefore, when synthesizing the findings from various studies, we took into account this variability in terminology and approach. For the purposes of this review, we adopted an inclusive approach and considered both types under the broader category of graphical knowledge mapping tools.

The integration of mind maps in STEM education has been widely explored, yet several critical aspects remain under discussion [21]. While the existing studies highlight the potential benefits of mind maps for learning, engagement, and cognitive processing [21], variations in their implementation, format, and pedagogical role raise important questions that require further investigation to understand their effectiveness and applicability across

different learning environments. For instance, the research comparing digital and paper-based mind maps is still limited, with some studies noting differences in flexibility and user experiences [22]. Similarly, the role of mind maps in collaborative versus individual learning contexts has yet to be systematically evaluated, although the evidence suggests that collaborative mind mapping may offer distinct advantages in fostering deeper engagement and higher-quality student outcomes [23].

To guide this systematic review, the following research questions were formulated:

1. What are the primary applications of mind maps in STEM education? Are they predominantly utilized as a learning tool or as an assessment instrument?
2. What research designs are most commonly employed to evaluate the effectiveness of mind maps in STEM education?
3. To what extent are mind maps implemented in individual learning contexts versus collaborative learning environments?
4. Are mind maps in STEM education primarily student-created or are they provided as ready-made resources?
5. What is the predominant format of mind maps in STEM education—paper-based or digital—and what implications does the format have for their effectiveness?
6. What learning outcomes have been reported in studies examining the use of mind maps in STEM education?

By addressing these questions, this review aims to provide a structured synthesis of the existing literature on mind maps in STEM education, highlighting trends, gaps, and areas for future research.

2. Materials and Methods

To identify pertinent studies on the use of mind maps in STEM education, we conducted a comprehensive search in three primary databases: ERIC, Scopus, and Web of Science. ERIC (Education Resources Information Center) was included for its extensive coverage of research and practice in education, making it a crucial source for pedagogical studies [24]. Scopus was selected for its broad, multidisciplinary scope and up-to-date indexing of peer-reviewed journals across the sciences and social sciences. Web of Science complemented these databases with its robust citation-tracking features and coverage of high-impact academic publications in STEM fields [25]. By drawing on these complementary resources, we aimed to capture both the foundational and emerging research on mind maps, ensuring that our review encompassed a diverse range of study designs, educational contexts, and disciplinary perspectives.

The systematic search and selection of studies adhered to the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [10]. This framework provided a structured, transparent process for identifying, screening, determining the eligibility of, and ultimately including articles in the review. By following the PRISMA methodology, we aimed to minimize bias and ensure a replicable and rigorous selection process. Each stage—ranging from the initial database searches to the final inclusion of studies—was documented using a PRISMA flow diagram [26] (Figure 1), offering clarity on how many records were retrieved, excluded, and retained for our in-depth analysis.

Inclusion Criteria

- Use of Concept or Mind Maps in STEM Education
  The study must explicitly examine concept maps, mind maps, or similar visual learning tools within a STEM (science, technology, engineering, or mathematics) context.
- Empirical Study with Measurable Outcomes
  The research must involve quantitative or qualitative data collection (e.g., pre–post-

tests, surveys, student performance metrics), demonstrating the impact of mind mapping tools on learning or related outcomes.

- Evaluation of Learning, Engagement, or Cognitive Processes
  The study must report at least one measurable aspect such as the learning performance, cognitive or metacognitive effects, engagement, or motivation.
- Recent Peer-Reviewed Publications
  Only articles published in peer-reviewed journals between 2019 and 2024 are included. Conference papers and other forms of grey literature are excluded.
- Relevant Educational Levels
  Studies covering primary, secondary, or higher education settings are included. Studies on teacher training in STEM are also considered if they measure student learning outcomes.

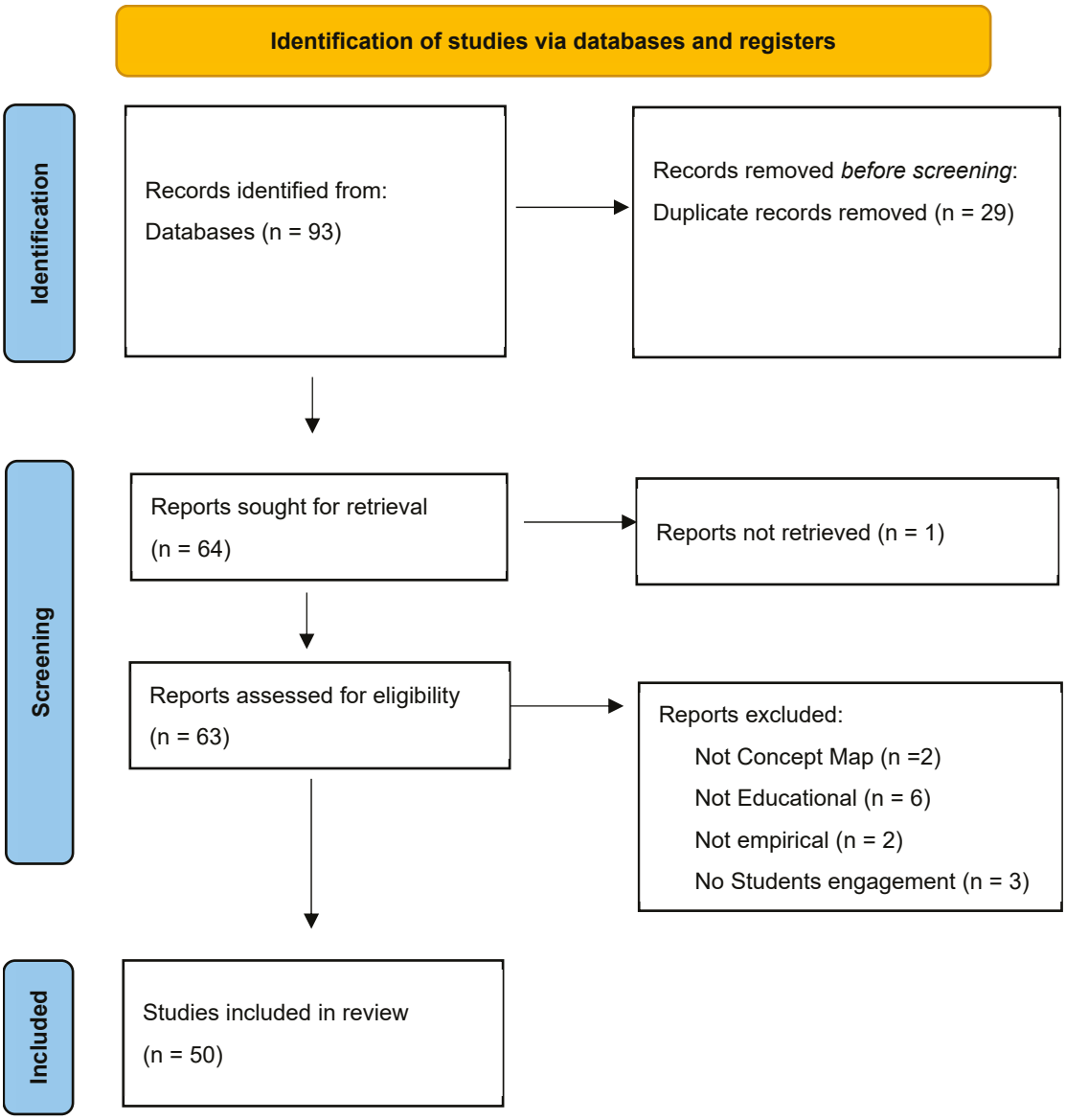


Figure 1. PRISMA flow diagram.

Exclusion Criteria

- No Direct Use of Concept or Mind Maps
  Studies focusing on other digital tools, visualization methods, or unrelated learning techniques without concept/mind maps as a core component are excluded.

- Non-STEM Focus
  Articles that do not address science, technology, engineering, or mathematics education are excluded.
- Professional or Workforce Training Context
  Studies centered on professional development, workforce training, or corporate learning contexts rather than formal education are excluded.
- Lack of Empirical Data
  Opinion pieces, theoretical papers, systematic reviews, or any study that does not report measurable learning or engagement outcomes is excluded.
- Duplicates/Redundant Records
  Any duplicate entries identified across the searched databases are excluded from the final dataset.

All search results from ERIC, Scopus, and Web of Science were imported into Mendeley, which served as the primary tool for managing references. Duplicate citations identified within Mendeley were systematically removed before proceeding with the screening process. In the first screening phase, titles and abstracts were reviewed against the predefined inclusion and exclusion criteria (e.g., focus on mind mapping in STEM education, empirical data). Articles that clearly did not meet these criteria were discarded. Potentially eligible studies were then retrieved in full text for a second, more detailed review. During this stage, any lingering uncertainties regarding a study's eligibility were resolved through discussion among the research team. This two-step approach ensured that only articles conforming to the specified criteria remained in the final pool for the data extraction and analysis.

A structured spreadsheet was used to document key information from each included study. For every article, we recorded basic metadata—such as the title, educational level, STEM subject, publication year, participants, and journal name—to establish the context. We then captured the primary use case (e.g., classroom instruction, assessment), the evaluation methodology (e.g., experimental, quasi-experimental, qualitative case study), and whether the study reported on cognitive load or other theoretical underpinnings. Additionally, we classified the intended outcomes as either learning-focused (e.g., facilitating comprehension, improving performance) or assessment-focused (e.g., evaluating student understanding through mind mapping). Finally, the key findings were noted to facilitate their later synthesis.

With regard to the mind-mapping interventions themselves, the mind map type indicated whether the mind map was pre-made (supplied by the instructor or software) or student-generated. The mind map format captured whether the mind maps were produced digitally (using specialized software) or on paper, while the interaction type distinguished between individual versus collaborative activities. This approach ensured that each study's unique characteristics and instructional strategies were documented consistently.

To maintain consistency, one reviewer extracted data from all included studies, while a second reviewer randomly sampled a subset of articles to verify the accuracy of the extracted information. Any discrepancies identified were resolved by re-examining the original full text and discussing the differences until consensus was reached. By recording these variables systematically, we established a robust foundation for analyzing patterns, trends, and gaps within the existing literature on mind maps in STEM education.

Quality Appraisal and Risk of Bias Assessment

To assess the methodological quality and internal validity of the included studies, a quality appraisal process was conducted using the Mixed Methods Appraisal Tool (MMAT, version 2018). This tool was selected due to its suitability for evaluating diverse empirical research designs, including quantitative (randomized and non-randomized), qualitative,

and mixed-methods studies [27]. Each included article was classified according to its methodological design and appraised using the corresponding set of five MMAT criteria.

For every study, a score was calculated by counting the number of criteria that were clearly met (e.g., 4/5 or 5/5), and a corresponding level of risk of bias was assigned. Studies that met all five criteria were classified as having low risk of bias, those with three or four met criteria were marked as having moderate risk, and those meeting two or fewer were considered to have high risk, following the guidance outlined in recent systematic reviews using the MMAT framework [28]. In cases where the design was descriptive or not explicitly stated, the evaluation was adapted using the MMAT's descriptive study criteria. All appraisals were independently verified by a second reviewer (author 2) to ensure consistency, and disagreements were resolved through discussion.

The appraisal results, including the MMAT score and overall risk of bias judgment for each study, are summarized in Appendix A (Table A1). This assessment provides transparency in evaluating the methodological rigor of the included studies and serves as a foundation for interpreting the reliability of the synthesized findings.

To integrate findings across the diverse set of studies, we employed a narrative synthesis approach. After data extraction, we organized the included articles based on major categories—such as the mind map type (pre-made vs. student-generated), mind map format (digital vs. paper-based), educational level, and intended outcomes (learning vs. assessment). This categorization allowed us to compare studies that shared similar characteristics, highlighting recurring patterns and distinct differences in their reported approaches and findings.

3. Results

This section presents the findings of the systematic review on the use of mind maps in STEM education. The analysis focused on the primary applications of mind maps, the research designs employed, the learning contexts in which they are implemented, the predominant formats used, and the reported learning outcomes. The reviewed studies span various educational levels and STEM disciplines, providing a comprehensive overview of how mind maps contribute to teaching and learning in these fields.

To ensure a structured synthesis, the findings are organized according to the research questions guiding this review. Initially, a summary table is provided (see Appendix A, Table A1), outlining key characteristics of the analyzed studies, including their objectives, methodologies, and main conclusions along with the MMAT scores and the corresponding risk of bias assessments. Following this, a more detailed discussion explores the trends and patterns identified across the selected studies.

The overview of the included studies is presented in Appendix A (Table A1).

3.1. Primary Applications of Mind Maps in STEM Education

A key objective of this review was to determine whether mind maps are predominantly utilized as learning tools or assessment instruments in STEM education. The analysis of the selected studies revealed that mind maps are primarily employed as learning tools (n = 40), whereas a smaller subset (n = 10) focused on their use as assessment instruments (Figure 2). Representative examples include Menouer et al. [6], Veiga et al. [29], and Petrun Sayers et al. [18] for learning tools, and Evans and Jeong [12] and Foley et al. [30] for assessment tools.

The majority of studies emphasize the role of mind maps in supporting knowledge organization, conceptual understanding, and problem-solving skills [6,9,29]. Mind maps serve as visual aids that enable students to structure information, identify relationships between concepts, and enhance their comprehension of complex STEM topics [10,30].

Several studies highlight their effectiveness in fostering active learning, promoting student engagement, and facilitating metacognitive reflection [14,16,31].

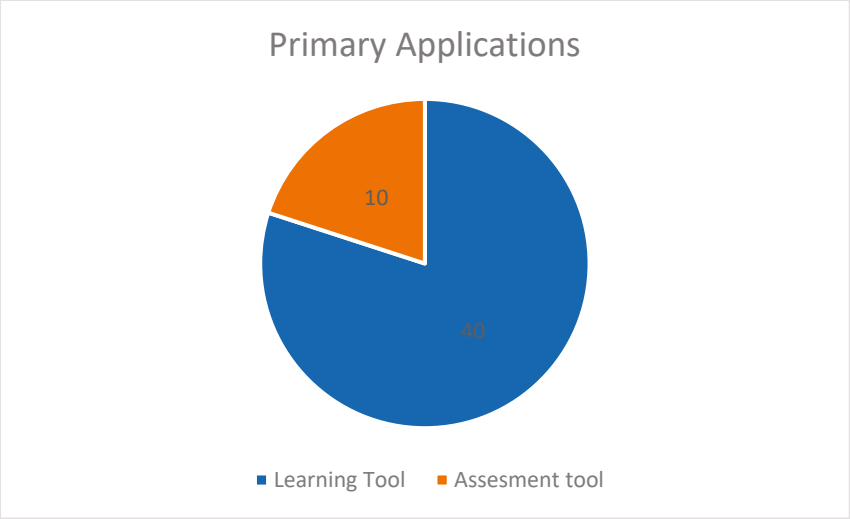


Figure 2. Primary applications of mind maps in STEM education. Based on data from the included studies. For detailed article-level categorization, see Appendix A, Table A1.

A smaller but significant body of research examines the use of mind maps as assessment tools. These studies demonstrate that concept mapping can serve as an alternative assessment method for evaluating student comprehension and cognitive structures [11,12]. Some research studies suggest that mind maps provide insights into students' thought processes and misconceptions, making them valuable for formative assessments [11,32].

3.2. Research Designs for Evaluating the Effectiveness of Mind Maps in STEM Education

The reviewed studies employed a variety of research designs to investigate the impact of mind maps in STEM education. The most prevalent approach was quasi-experimental research (n = 29), as seen in studies such as those by Menouer et al. [6], Fang et al. [14], and Hwang et al. [16], with the majority (n = 22) incorporating pre- and post-test measurements to assess learning gains. This suggests a strong emphasis on quantifying the effectiveness of mind maps in improving student outcomes. Additionally, case studies (n = 8) were also employed, including in the studies by Sagarika an Syedkhamruddin [3] and Bredeweg et al. [17], allowing for in-depth qualitative insights. Mixed-methods research (n = 9), such as in the studies by Veiga et al. [29] and Chen et al. [33], combines qualitative and quantitative approaches to capture broader outcomes. A smaller number of studies (n = 4) adopted descriptive designs, such as for Waghmare et al. [8] and Dewi et al. [34], focusing on observational insights without intervention.

The widespread use of quasi-experimental and mixed-methods approaches highlights a growing trend toward evidence-based evaluations of mind maps' effectiveness in STEM education (Figure 3). The full distribution of the studies by research design is provided in Appendix A, Table A1.

3.3. Individual vs. Collaborative Learning Contexts

The implementation of mind maps in STEM education varies between individual and collaborative learning contexts. Mind maps are most commonly used in individual learning settings (n = 24), as seen in the studies by Menouer et al. [6], Debbag et al. [22], and Jeong and Evans [10], where students independently constructed mind maps to organize knowledge and for improved understanding. Collaborative applications (n = 11) were employed in studies such as those by Krab-Hüsken et al. [9] and Fung and Liang [35],

highlighting the group co-construction of maps to foster peer interaction and problem-solving. A mixed approach (n = 15), combining both individual and group activities, was found in the studies by Hwang et al. [16] and Susithra et al. [11], where students began mapping individually and later collaborated to refine their ideas. Overall, while individual use remains the dominant approach, the increasing application of collaborative and mixed methods suggests a shift toward more interactive and socially engaged learning experiences in STEM education (Figure 4). The complete list of studies by learning context is presented in Appendix A, Table A1.

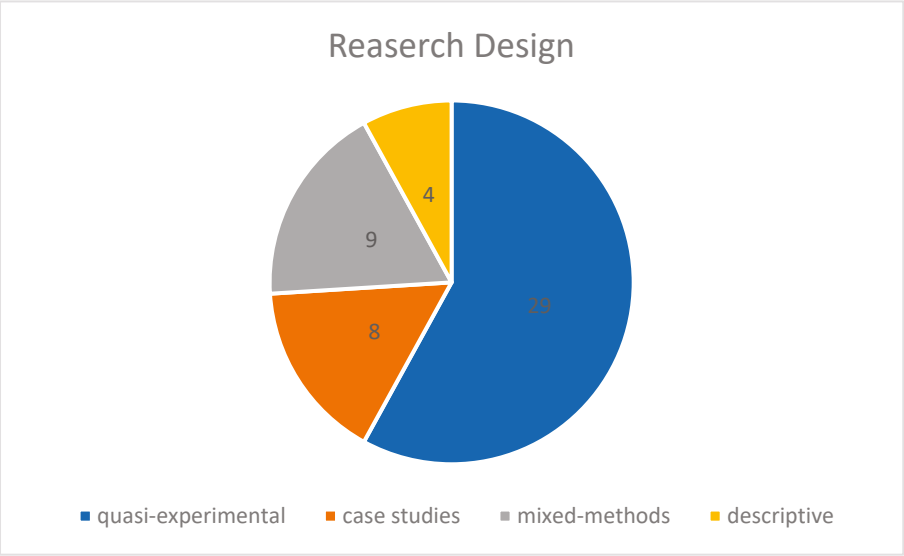


Figure 3. Research designs used in the reviewed studies. Based on data from the included studies. For detailed article-level categorization, see Appendix A, Table A1.

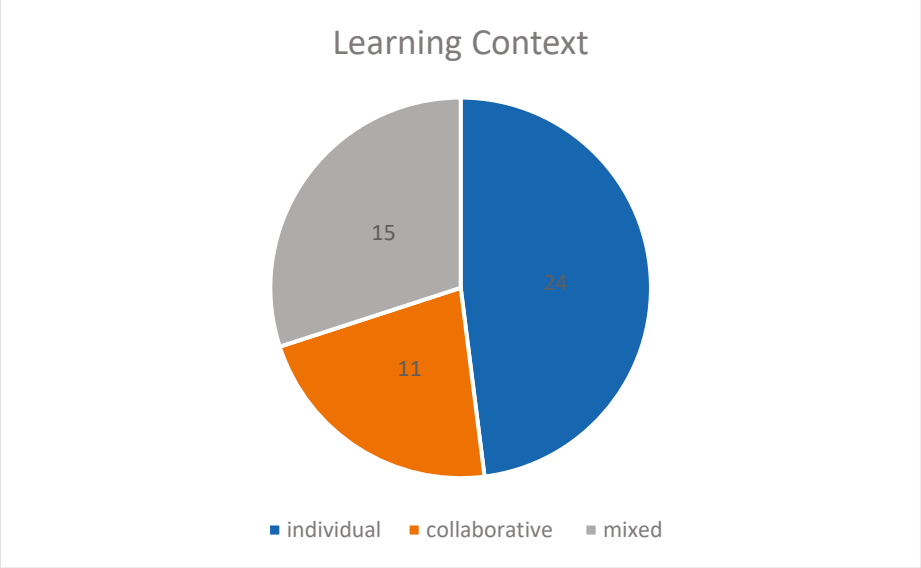


Figure 4. Learning context.

3.4. Student-Created vs. Ready-Made Mind Maps

An important consideration in the use of mind maps in STEM education is whether they are constructed by students or provided as ready-made resources. The findings indicate that student-created mind maps are the predominant approach (n = 36), as illustrated in studies such as those by Fang et al. [14], Chen and Chung [4], and Astriani et al. [13], suggesting a strong emphasis on active engagement, knowledge construction, and self-

regulated learning. In contrast, ready-made mind maps (n = 9) in the studies by Debbag et al. [22] and Krab-Hüsken et al. [9] were typically used to scaffold instructions and provide pre-structured knowledge frameworks. Additionally, 5 studies did not explicitly state whether their mind maps were student-created or ready-made, indicating a gap in the reporting of this methodological aspect. The full classification of the studies based on map origin is available in Appendix A, Table A1.

Student-generated mind maps are often associated with higher cognitive engagement, as they require learners to actively organize information and establish connections between concepts. On the other hand, ready-made mind maps can be beneficial for scaffolding learning, particularly when introducing complex topics or supporting students with limited prior knowledge. The choice between these two approaches may depend on the learning objectives, student proficiency levels, and instructional design considerations (Figure 5). This aligns with previous research highlighting the importance of adapting digital learning tools, such as mind maps, to support the diverse needs of students, including gifted learners [36].

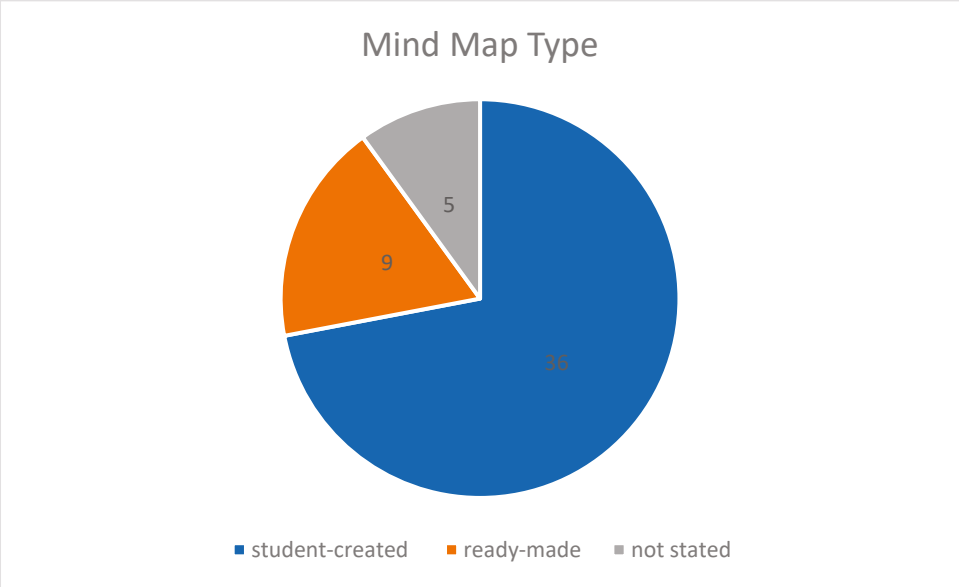


Figure 5. Student-created vs. ready-made mind maps, based on data from the included studies. For a detailed article-level categorization, see Appendix A, Table A1.

3.5. Predominant Format of Mind Maps

The format of mind maps—whether paper-based or digital—can influence their effectiveness in STEM education. The findings indicate that digital mind maps are slightly more prevalent (n = 21), as shown in studies such as those by Chen and Chung [4], Hwang et al. [16], and Veiga et al. [29], where digital tools supported interactivity and multimedia integration. Paper-based formats (n = 18) were used in the studies by Krab-Hüsken et al. [9] and Jeong and Evans [10], often favoring traditional sketching for conceptual organization and spatial memory retention. Some studies (n = 4) combined both formats, as in the study by Debbag et al. [22], while 9 studies did not specify the format employed (Figure 6). Digital mind maps offer advantages in flexibility, interactivity, and ease of modification, making them particularly useful for dynamic learning environments and collaborative activities [4,16,29]. Many digital tools also integrate features such as hyperlinks, multimedia elements, and automatic structuring, which can enhance students' engagement and comprehension [22]. On the other hand, paper-based mind maps are still widely used, particularly in traditional classroom settings where students sketch out their ideas manually [9,10]. Some research suggests that the physical act of drawing mind maps

may strengthen students' cognitive processing and support spatial memory retention [10]. However, paper-based maps lack the ease of revision and organization offered by digital alternatives [22].

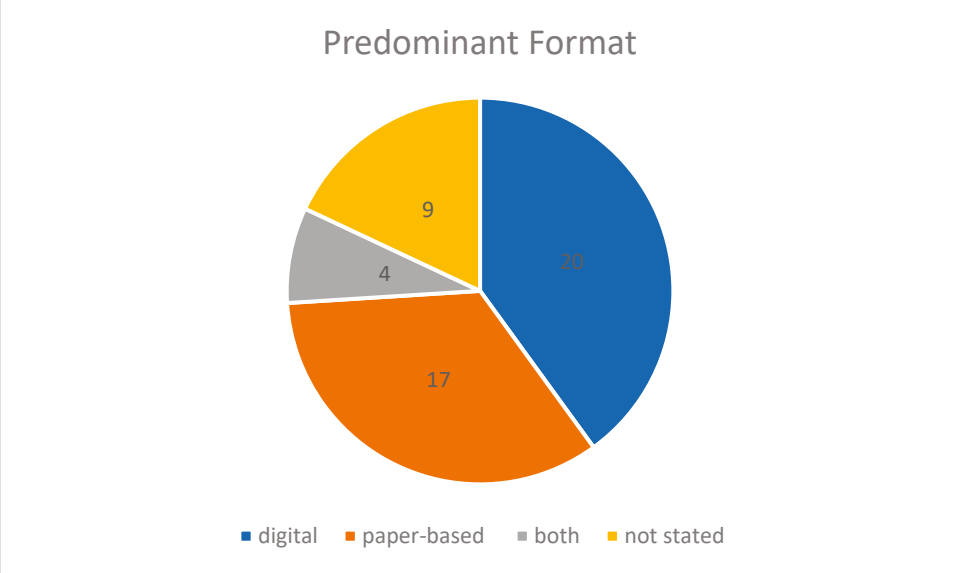


Figure 6. Predominant format of mind maps: digital vs. paper-based. Based on data from the included studies. For a detailed article-level categorization, see Appendix A, Table A1.

The studies that used both formats ($n = 4$) indicate that hybrid approaches may provide the best of both worlds, allowing students to brainstorm on paper before transferring ideas into digital tools for refinement and sharing [22].

3.6. Learning Outcomes in Mind Mapping for STEM Education

In the studies we examined, various learning outcomes were reported as a result of using mind maps in STEM education. These outcomes were categorized to help better understand the impacts of mind maps on student learning and achievement. The categories are discussed below.

3.6.1. Academic Performance and Problem-Solving

The use of concept maps in STEM education has been shown to significantly improve students' academic performance and problem-solving abilities [6,30]. Studies report notable improvements in test scores, problem-solving efficiency, and engagement. For instance, students' analysis and conclusion skills were significantly enhanced, with their reasoning skills showing marked improvements [6]. Concept maps also contributed to more complex individual maps, which in turn positively influenced students' academic performance [29]. Furthermore, concept mapping reduced students' textbook reading time and boosted their problem-solving efficiency, with the experimental group scoring significantly higher in the post-tests [37]. Students who used concept maps also demonstrated better end-of-semester exam performance and found the tool effective for visualizing complex topics and connecting prior knowledge with new concepts [30]. In engineering education, concept maps helped students integrate social and technical dimensions of sociotechnical systems, making them a valuable tool for assessing learning outcomes [3]. Additionally, in the context of computational thinking, concept maps helped students structure programming logic, debug errors, and increase engagement [4]. Knowledge mapping also supported students' personalized learning and competence development, although the initial cognitive load was higher for unfamiliar students [38].

3.6.2. Conceptual Understanding and Knowledge Organization

Concept maps in STEM education have been shown to significantly enhance students' conceptual understanding and their ability to organize knowledge [8,10]. Studies have highlighted that concept maps help students visually organize and connect ideas, fostering clarity, creativity, and accuracy in their learning [8]. Concept maps were particularly effective in representing complex systems, such as sustainability systems, and helped students develop systems thinking by linking various elements such as leverage points and delays [1]. However, not all students were able to produce structured maps; while high-achieving students showed precise language use, they did not always create well-structured sensemaking maps, and contextual factors played a role in these outcomes [2]. The use of concept maps also improved students' understanding of ecology concepts, promoting a shift from surface learning to a deeper understanding [39]. Over time, the concept maps have grown more integrated, with abstract terms becoming more central, reflecting a nonlinear process of conceptual change [40]. Concept maps also facilitated deeper connections between mathematical concepts, helping students support relational reasoning and achieve higher scores by spending more time on tasks [10]. The alignment between formative and summative assessments showed that mind maps effectively enhanced students' ability to connect and integrate concepts for exam preparation [11]. In sustainability education, cognitive mapping improved the interconnectedness of sustainability concepts among STEM students, revealing significant improvements in conceptual complexity [18]. Moreover, design-centric courses that incorporated concept mapping fostered more transformative learning experiences, although the students needed further engagement beyond conceptual mastery for real-world application [33].

3.6.3. Critical Thinking and Higher-Order Skills

Mind mapping has proven to be an effective tool for enhancing critical thinking and higher-order thinking skills in STEM education [41,42]. Studies have indicated that concept maps positively impacted not only critical thinking but also motivation and teamwork skills [43]. Additionally, mind mapping was linked to significant improvements in metacognitive skills, such as the ability to plan, monitor, and evaluate learning processes, with post-test scores showing substantial gains [13]. These findings are in line with metacognitive models highlighting the role of visual tools such as mind maps in enhancing memory operations and self-regulated learning strategies through digital environments [44]. The use of concept maps also helped students engage more deeply in problem-solving and critical thinking, especially in group settings, where scaffolding from concept maps enhanced students' learning outcomes and reduced problem-posing challenges [16]. In [41], the use of concept mapping through the KC-CMG strategy improved students' learning achievement, problem-solving awareness, and critical thinking, with the students showing a greater tendency to revisit concept maps after mistakes, which in turn improved their motivation and reflection. Furthermore, concept mapping facilitated the development of higher-order thinking skills in structured debates and a healthcare argument synthesis, promoting a transition from passive to active learning [42]. Although some students initially struggled with structuring their arguments, the overall cognitive load was reduced over time, allowing the students to engage more effectively in higher-order thinking tasks.

3.6.4. Motivation, Engagement, and Self-Efficacy

Mind mapping has been found to significantly enhance motivation, engagement, and self-efficacy in STEM education [12,45]. Several studies have demonstrated that concept maps not only improve academic outcomes but also boost students' emotional resilience and confidence. One study showed that concept mapping explained a significant vari-

ance in final exam scores, while also being linked to emotional regulation, suggesting its potential to enhance students' self-efficacy during learning [12]. Other tools, such as DynaLearn, Minds-On, and ICC, were also found to enhance learner engagement and students' conceptual understanding, particularly in systems thinking, although further scaffolding was needed for optimal use [17]. Additionally, students using mind maps showed higher concept understanding and better retention, with increased motivation and confidence in exams, indicating that mind mapping served as an effective study aid [46]. Another study found significant improvements in comprehension scores, along with positive feedback from parents on increased interest and engagement in reading, further demonstrating the effectiveness of mind maps in motivating students [47]. Further evidence revealed that concept mapping significantly boosted self-efficacy and self-regulation, with students displaying better goal-setting and time management skills, as well as increased participation and motivation in STEM projects [14]. Another study showed that DIME maps had a strong positive impact on self-efficacy, with female students particularly benefiting, although some found the maps overwhelming initially due to their information density [48]. Finally, knowledge mapping was found to support personalized learning and competence development, leading to better project performance, although the initial cognitive load was high for unfamiliar students [49]. A previous study also highlighted how students associated STEM with anxiety and negativity, indicating the potential of mindset-focused interventions to improve engagement with STEM [45].

3.6.5. Cognitive Load and Information Processing

Mind mapping has been shown to significantly affect the cognitive load and information processing in STEM education [50]. Several studies have indicated that concept maps help reduce the cognitive load by organizing complex ideas and improving information retrieval. For example, concept mapping was found to reduce the neurocognitive effort compared to traditional brainstorming, with students reporting higher satisfaction and effectiveness in organizing ideas and generating innovative solutions [51]. However, physical computing, unplugged computing, and PRIMM were highly rated for autistic students due to their emphasis on hands-on and visual learning, while concept maps ranked lower due to challenges in their implementation and perceived difficulty [52]. Another study highlighted the advantages of PBMM (physical-based mind mapping) for psychomotor development and collaborative learning, while DMM (digital mind mapping) was preferred for its multimedia capabilities, such as easy editing and the addition of images and videos [22]. The study also found that combining methods such as biodrawing followed by mind mapping helped to maximize the learning outcomes, even though the mind maps were valuable mainly for early-stage brainstorming [34]. In terms of learning achievement, one study revealed that ARG-CMQ (augmented reality game-based concept mapping) significantly improved students' learning outcomes in single-player phases, although it led to a higher cognitive load and struggles with complex information processing in multiplayer phases [53]. Similarly, in another study, students reported that concept maps helped identify knowledge gaps and improve their problem-solving approaches, although there was no significant impact on the final performance [31]. Concept maps improved students' navigation coherence and helped them engage more deeply in the learning process but they could also be overwhelming for some students, especially early on [54]. In collaborative learning settings, concept mapping helped improve students' conceptual understanding, although structured guidance was needed to reduce the cognitive overload. The combination of augmented reality (AR) and concept mapping also improved students' learning performance, particularly for those students with low prior knowledge, as it helped to reduce the cognitive load and improve their knowledge retention [55]. Furthermore, in

virtual reality (VR) contexts, concept maps enhanced students' learning achievement and reflection, although some students preferred more interactive feedback mechanisms [50]. Another study showed that higher-quality feedback was a stronger predictor of learning success compared to the product type, such as concept mapping [32]. Concept mapping also facilitated a shift from surface learning to deeper learning strategies, improving student engagement and reducing rote memorization [56]. Additionally, in science literacy, synthesis maps were found to improve learning outcomes, with students performing better in journal club presentations by structuring research narratives with improved logical flow [57]. Finally, concept mapping training was shown to improve long-term retention and reduce the cognitive load, with students outperforming control groups and showing enhanced organizational skills and error detection accuracy [58]. Lastly, concept cards were found to enhance the concept mapping performance by helping students recall concepts faster and making learning more engaging, especially for low-scoring students [59].

3.6.6. Collaboration and Communication Skills

Mind mapping has been shown to significantly enhance collaboration and communication skills in STEM education [9,60]. Studies have indicated that concept maps support conceptual modeling, which helps students structure their ideas and promotes systems thinking. This also led to improved collaboration in group projects, fostering engagement and openness to diverse perspectives, with students recommending additional scaffolding and feedback to support the process [9]. In cross-border discussions, concept maps were used to effectively summarize ideas, capture connections, and facilitate collaboration, fostering critical thinking. The cross-border format revealed gaps in the conceptual understanding, such as proportional reasoning, which were identified during collaborative learning [61]. Furthermore, high-performing groups were more effective in divergent thinking strategies and regulative discussions, which led to enhanced idea generation and improved collaborative discussions. In contrast, lower-performing groups struggled with integrating divergent thinking and mind mapping [60]. The integration of concept mapping into STEM projects led to significant improvements in self-efficacy, self-regulation, and project outcomes, helping students develop better goal-setting, task strategy, time management, and reflection skills, all of which contributed to increased participation and motivation [35]. In computational contexts, students using concept maps demonstrated better problem-solving and computational thinking, with higher engagement reported [62]. Finally, collaborative mind mapping was found to improve organization and engagement, particularly in mathematics education. It enhanced students' information retention and critical thinking, although some students experienced cognitive overload when organizing large datasets. The study suggested that students engaged in the continued use of mind maps would benefit from further training in structuring such maps effectively [63].

Figure 6 visually represents the key outcomes identified across the studies, providing a clearer understanding of how mind maps influence different aspects of student performance and skills in STEM education. In order to provide transparency regarding the distribution of studies across the six outcome categories presented in Figure 7, Table 1 below lists the specific articles associated with each category.

Table 1. Mapping of reviewed studies to learning outcome categories.

Learning Outcome Category	Studies
Academic Performance and Problem-Solving	[3,4,6,29,30,37,38]
Conceptual Understanding and Knowledge Organization	[1,2,8,10,11,18,33,39,40]
Critical Thinking and Higher-Order Thinking Skills	[13,16,41–43]

Table 1. *Cont.*

Learning Outcome Category	Studies
Motivation, Engagement, and Self-Efficacy	[12,14,17,45–49]
Cognitive Load and Information Processing	[22,23,31,32,34,50–58],
Collaboration and Communication Skills	[9,35,60–63]

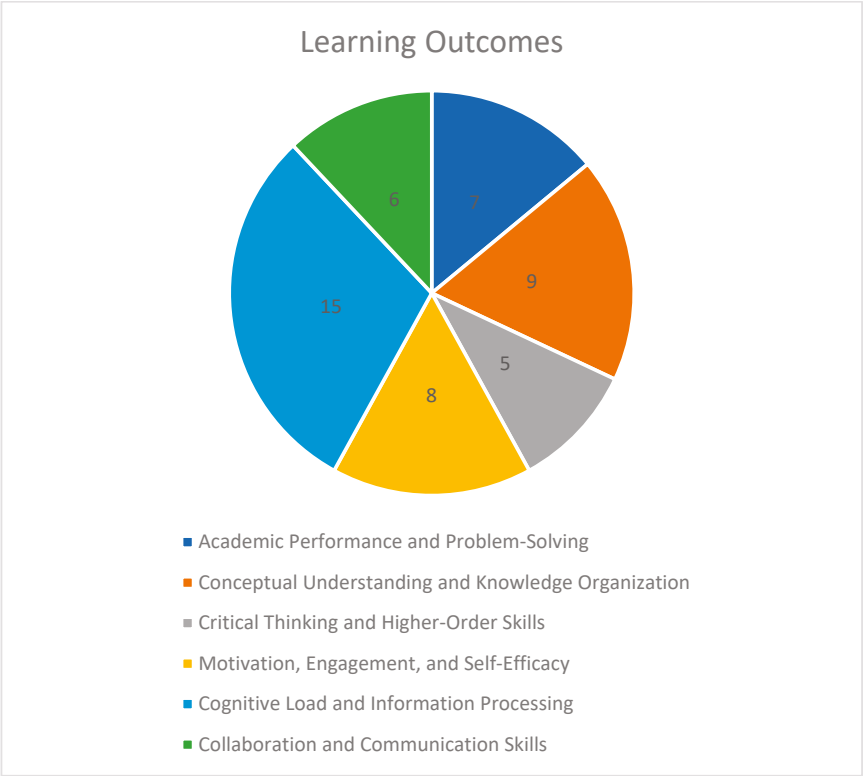


Figure 7. Learning outcomes. For a breakdown of the studies included in each category, see Table 1 below.

4. Discussion

The validity and generalizability of the findings in this review are shaped by the methodological quality of the included studies. As indicated in the quality appraisal using the MMAT (2018), the majority of studies (n = 31) were rated as having low risk of bias, while 18 studies were classified as moderate-risk and 1 as high-risk (Appendix A, Table A1). Although this distribution suggests that the overall evidence base is relatively strong, the presence of studies with moderate or high risk of bias—particularly those lacking control groups, using small samples, or providing limited methodological transparency—may affect the robustness of some conclusions. For example, several descriptive studies and case studies did not clearly specify the type or format of the mind maps used, limiting their replicability. Consequently, while the findings highlight the positive impacts of mind maps on learning and engagement in STEM education, they should be interpreted with caution. Further research employing more rigorous experimental designs and transparent reporting is needed to enhance the reliability and transferability of the results across different educational contexts.

The results of this systematic review on the use of mind maps in STEM education reveal several important findings regarding their application and effectiveness and the various factors that influence their use in educational settings. The following discussion will

provide an interpretation of these results, highlighting their implications for educational practice, and suggesting directions for future research.

The findings show that mind maps are predominantly employed as learning tools rather than assessment instruments. This is evident in studies such as those Menouer et al. [6] and Fang et al. [14], where mind maps were used to support knowledge construction and problem-solving. The widespread use of mind maps in individual learning contexts emphasizes their role in organizing knowledge, improving one's conceptual understanding, and supporting problem-solving skills [8,10]. However, the smaller subset of studies using mind maps as assessment tools, including those by Evans and Jeong [12] and Susithra et al. [11], indicate their potential in evaluating students' cognitive structures and comprehension. These findings contribute to the growing body of research suggesting that concept mapping can serve as a formative assessment tool, providing insights into students' thought processes and misconceptions [31,50].

The majority of studies employed quasi-experimental designs, often incorporating pre- and post-test measurements, as shown in the studies by Huang et al. [30] and Hwang et al. [16], suggesting a strong emphasis on measuring the effectiveness of mind maps in improving learning outcomes. The use of case studies (e.g., Sagarika and Syedkhamruddin [3]; Bredeweg et al. [17]) and mixed-methods designs (e.g., Veiga et al. [29]; Chen et al. [38]) further highlights the value of combining quantitative and qualitative data. The predominance of quasi-experimental designs reinforces the importance of using evidence-based approaches in educational research. However, several studies call for more robust longitudinal investigations to better explore the long-term effects of mind maps on student learning outcomes and retention [32,57].

The results indicate that student-created mind maps are the dominant approach, supporting the view that active engagement in knowledge construction leads to higher cognitive engagement [4,14,38]. This aligns with the constructivist theory, which emphasizes the importance of learners actively building their own knowledge. On the other hand, ready-made mind maps serve as valuable scaffolds, particularly in studies such as those by Krab-Hüsken et al. [9] and Debbag et al. [22], which provided s structure for complex content or novice learners. The mixed approach reported in studies such as the one performed by Hwang et al. [16] suggests that a hybrid model might offer the best of both worlds, allowing students to begin with their own representations and refine them using pre-structured maps.

The results show a slight preference for digital mind maps, which offer greater flexibility and interactivity and the ability to integrate multimedia elements. These findings reflect the growing trend of using technology-supported tools in education, which aligns with previous research that emphasizes the advantages of digital tools in enhancing students' engagement and conceptual understanding [7,36,64]. However, the continued use of paper-based maps suggests that traditional tools still play a significant role in education, particularly in settings where digital access is limited. Studies that use both formats indicate the potential of a hybrid approach, where students begin with paper-based maps to facilitate brainstorming and then transfer their ideas into digital tools for further development and sharing. This hybrid model provides a balanced approach, combining the tactile benefits of paper with the organizational advantages of digital tools.

The results demonstrate that mind mapping significantly improves students' academic performance, as evidenced by improvements in test scores, exam performance, and problem-solving efficiency. These findings align with previous research that has shown the effectiveness of visual tools such as concept maps in helping students organize and retain complex information [3,37,38]. The positive correlation between concept map quality and learning outcomes suggests that these tools can be particularly effective for exam

preparation [11]. Furthermore, the reduction in textbook reading and increased problem-solving efficiency indicate that concept maps allow students to internalize knowledge more effectively, promoting deeper cognitive processing. These findings align with the principles of algorithmic and procedural learning, where students must follow structured steps and logical sequences to arrive at solutions. Mind maps can support this process by helping learners break down complex problems into manageable components and visualize dependencies between steps. In programming and mathematical contexts, such structured visual thinking is particularly beneficial for developing algorithmic fluency and procedural accuracy.

Future research should aim to investigate the long-term effects of mind mapping on learning retention through longitudinal studies, as well as explore its differentiated impacts across STEM disciplines. Comparative studies between digital and paper-based mind maps—especially in collaborative versus individual learning settings—could offer deeper insights into their context-specific effectiveness. In addition, the development of standardized evaluation tools for measuring cognitive, metacognitive, and affective outcomes in mind-map-based interventions would enhance the consistency and comparability across studies.

Limitations

This review combined studies using both concept maps and mind maps, despite their acknowledged structural and methodological differences. While this reflects common usage patterns in the reviewed literature, it may have introduced some variability in the synthesis of the findings. Future research could benefit from explicitly distinguishing between these tools to further clarify their specific impacts on learning outcomes.

5. Conclusions

This systematic review highlights the multifaceted role of mind maps in STEM education, emphasizing their effectiveness as both learning and assessment tools. The majority of studies demonstrate that mind maps significantly enhance students' academic performance, conceptual understanding, critical thinking, and engagement. Student-generated maps, in particular, appear to support active learning and cognitive development, while digital formats offer greater interactivity and adaptability to modern educational contexts. Furthermore, mind mapping has been associated with increased motivation and self-efficacy, especially when integrated into collaborative or mixed learning environments. While most studies reported reductions in cognitive load, some noted that unfamiliarity with the technique or excessive complexity could cause temporarily increased mental effort. Despite these positive outcomes, the review also reveals gaps in the literature, particularly regarding the long-term effects of mind mapping, its use across different STEM subjects, and the comparative impacts of digital versus paper-based formats. Future research should aim to standardize the assessment approaches, explore the scalability of mind maps across diverse educational settings, and further examine their role in promoting self-regulated and interdisciplinary learning. Mind maps offer a promising, research-supported strategy for enriching STEM education. Their continued integration—guided by empirical evidence and thoughtful instructional design—can support deeper, more connected, and more personalized learning experiences for students at all levels.

Author Contributions: Conceptualization, C.K., C.S. and A.D.; methodology, C.K., C.S. and A.D.; validation, C.K., C.S. and A.D.; formal analysis, C.K., C.S. and A.D.; investigation, C.K., C.S. and A.D.; resources, C.K., C.S. and A.D.; writing—original draft preparation, C.K., C.S. and A.D.; writing—review and editing, C.K., C.S. and A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

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

Appendix A

Table A1. Quality appraisal of included studies using the Mixed Methods Appraisal Tool (MMAT, version 2018). The MMAT score (out of 5) reflects the number of criteria met for the relevant study design. The risk of bias was categorized as low, moderate, or high based on scoring thresholds and methodological transparency.

Title	Research Design	Primary Applications	Learning Context	Predominant Format	Score (Yes/Total)	Quality (Risk of Bias)
Effects of Employing Mind Mapping Techniques in Geometry Instruction on Logical Thinking Abilities [6]	Quasi-Experimental	Learning Tool	Individual	Not Stated	5/5	Low
The Significance of Technology-Enhanced Learning towards Enhancing Engineering Education [8]	Descriptive	Learning Tool	Individual	Digital	3/5	Moderate
Lessons for Sustainable Science Education: A Study on Chemists' Use of Systems Thinking across Ecological, Economic, and Social Domains [1]	Quasi-Experimental	Learning Tool	Individual	Paper-Based	4/5	Moderate
The Role of Scientific Language Use and Achievement Level in Student Sensemaking [2]	Mixed Methods	Learning Tool	Individual	Not Stated	5/5	Low
Investigating The Transformative Effects of Active Learning Methodologies in The Field of Engineering Education to Improve Learning Outcomes in Students by Unleashing Their Potential [43]	Quasi-Experimental	Learning Tool	Mixed	Not Stated	5/5	Low
Validation of the Use of Concept Maps as an Evaluation Tool for the Teaching and Learning of Mechanical and Industrial Engineering [29]	Mixed Methods	Learning Tool	Mixed	Both	5/5	Low
Optimizing Inquiry-Based Science Education: Verifying the Learning Effectiveness of Augmented Reality and Concept Mapping in Elementary School [37]	Quasi-Experimental	Learning Tool	Collaborative	Not Stated	5/5	Low
Concept Maps as Assessment for Learning in University Mathematics [12]	Quasi-Experimental	Assessment Tool	Individual	Digital	5/5	Low
Learning with Interactive Knowledge Representations [17]	Case Study	Learning Tool	Mixed	Digital	5/5	Low
Square Pegs and Round Holes: Pedagogy for Autistic Students in Computing Education [52]	Descriptive	Learning Tool	Mixed	Not Stated	4/5	Moderate

Table A1. *Cont.*

Title	Research Design	Primary Applications	Learning Context	Predominant Format	Score (Yes/Total)	Quality (Risk of Bias)
Conceptual Modeling Enables Systems Thinking in Sustainable Chemistry and Chemical Engineering [9]	Quasi-Experimental	Learning Tool	Collaborative	Paper-Based	5/5	Low
Empirical Evidence That Concept Mapping Reduces Neurocognitive Effort During Concept Generation for Sustainability [51]	Mixed Methods	Learning Tool	Individual	Not Stated	5/5	Low
Using Immersive and Modelling Environments to Build Scientific Capacity in Primary Preservice Teacher Education [39]	Descriptive	Assessment Tool	Collaborative	Paper-Based	4/5	Moderate
A Novel Methodology for Improving Teaching Learning Process and Its Outcome on 2K Students for Engineering Education [46]	Mixed Methods	Learning Tool	Individual	Not Stated	5/5	Low
Mind Mapping in Learning Models: A Tool to Improve Student Metacognitive Skills [13]	Quasi-Experimental	Learning Tool	Individual	Not Stated	4/5	Moderate
Experiencing the Essence of Learning Database Management System Course Using C-map Tool [3]	Case Study	Learning Tool	Collaborative	Digital	3/5	High
Concept Mapping in Magnetism and Electrostatics: Core Concepts and Development over Time [40]	Quasi-Experimental	Assessment Tool	Individual	Digital	4/5	Moderate
Using a Concept Map With RECALL to Increase the Comprehension of Science Texts for Children With Autism [47]	Case Study	Learning Tool	Individual	Paper-Based	4/5	Moderate
Measuring the Amorphous: Substantive and Methodological Outcomes from Concept Maps [30]	Quasi-Experimental	Assessment Tool	Individual	Paper-Based	4/5	Moderate
Use of Digital Mind Maps in Technology Education: A Pilot Study with Pre-Service Science Teachers [22]	Case Study	Learning Tool	Mixed	Both	5/5	Low
Are Cross-Border Classes Feasible for Students to Collaborate in the Analysis of Energy Efficiency Strategies for Socioeconomic Development While Keeping CO2 Concentration Controlled? [61]	Quasi-Experimental	Learning Tool	Collaborative	Digital	4/5	Moderate
Powering Up Flipped Learning: An Online Learning Environment with a Concept Map-Guided Problem-Posing Strategy [16]	Quasi-Experimental	Learning Tool	Mixed	Digital	4/5	Moderate
Analysis of the Effectiveness of Architectural Creative Learning Methods [34]	Descriptive	Learning Tool	Individual	Paper-Based	4/5	Low
How Do Students Generate Ideas Together in Scientific Creativity Tasks Through Computer-Based Mind Mapping? [60]	Quasi-Experimental	Learning Tool	Collaborative	Digital	4/5	Moderate

Table A1. *Cont.*

Title	Research Design	Primary Applications	Learning Context	Predominant Format	Score (Yes/Total)	Quality (Risk of Bias)
Knowledge Organisers for Learning: Examples, Non-Examples and Concept Maps in University Mathematics [10]	Case Study	Learning Tool	Individual	Paper-Based	5/5	Low
Coalescing Mind Maps as a Learning Aid and Formative Assessment Tool for Effective Teaching and Learning of Computer Architecture and Organization Course [11]	Case Study	Learning Tool	Mixed	Paper-Based	5/5	Low
A Concept Mapping-Based Self-Regulated Learning Approach to Promoting Students' Learning Achievement and Self-Regulation in STEM Activities [14]	Quasi-Experimental	Learning Tool	Collaborative	Digital	5/5	Low
The Effectiveness of Collaborative Mind Mapping in Hong Kong Primary Science Classrooms [35]	Quasi-Experimental	Learning Tool	Collaborative	Paper-Based	5/5	Low
Fostering Computational Thinking and Problem-Solving in Programming: Integrating Concept Maps Into Robot Block-Based Programming [4]	Quasi-Experimental	Learning Tool	Individual	Digital	4/5	Moderate
Using Concept Maps to Analyze Educators' Conceptions of STEM Education [62]	Case Study	Learning Tool	Individual	Not Stated	5/5	Low
Knowledge Check-Based Concept Mapping in Digital Games: Impacts on Students' Learning Performance and Behaviors [41]	Quasi-Experimental	Learning Tool	Individual	Digital	5/5	Low
Promoting Children's Inquiry Performances in Alternate Reality Games: A Mobile Concept Mapping-Based Questioning Approach [53]	Quasi-Experimental	Learning Tool	Mixed	Digital	5/5	Low
Concept Mapping as a Metacognition Tool in a Problem-Solving-Based BME Course During In-Person and Online Instruction [31]	Quasi-Experimental	Assessment Tool	Mixed	Both	4/5	Moderate
Improving Self-Efficacy With Automatically Generated Interactive Concept Maps: DIME Maps [48]	Quasi-Experimental	Learning Tool	Individual	Digital	5/5	Low
Do Graphic and Textual Interactive Content Organizers Have the Same Impact on Hypertext Processing and Learning Outcomes? [65]	Quasi-Experimental	Learning Tool	Individual	Digital	5/5	Low
Molecular Orbital Theory—Teaching a Difficult Chemistry Topic Using a CSCL Approach in a First-Year University Course [54]	Quasi-Experimental	Learning Tool	Collaborative	Digital	4/5	Moderate

Table A1. *Cont.*

Title	Research Design	Primary Applications	Learning Context	Predominant Format	Score (Yes/Total)	Quality (Risk of Bias)
What Factors Influence Scientific Concept Learning? A Study Based on the Fuzzy-Set Qualitative Comparative Analysis [55]	Quasi-Experimental	Learning Tool	Collaborative	Paper-Based	5/5	Low
A Concept Map-Based Community of Inquiry Framework for Virtual Learning Contexts to Enhance Students' Earth Science Learning Achievement and Reflection Tendency [50]	Quasi-Experimental	Learning Tool	Mixed	Digital	5/5	Low
Does Learning from Giving Feedback Depend on the Product Being Reviewed: Concept Maps or Answers to Test Questions? [32]	Quasi-Experimental	Assessment Tool	Individual	Digital	5/5	Low
The Use of Activity, Classroom Discussion, and Exercise (ACE) Teaching Cycle for Improving Students' Engagement in Learning Elementary Linear Algebra [56]	Mixed Methods	Learning Tool	Mixed	Paper-Based	5/5	Low
Synthesizing Research Narratives to Reveal the Big Picture: A CREATE(S) Intervention Modified for Journal Club Improves Undergraduate Science Literacy [57]	Quasi-Experimental	Learning Tool	Mixed	Paper-Based	4/5	Moderate
Comprehension-Oriented Learning of Cell Biology: Do Different Training Conditions Affect Students' Learning Success Differentially? [58]	Quasi-Experimental	Learning Tool	Individual	Paper-Based	4/5	Moderate
Online Argumentation-Based Learning Aided by Digital Concept Mapping During COVID-19: Implications for Health Management Teaching and Learning [42]	Case Study	Learning Tool	Mixed	Digital	5/5	Low
Pedagogically-Informed Knowledge Mapping: Representing Contextualised Competences and Technology Implemented [49]	Quasi-Experimental	Learning Tool	Mixed	Digital	4/5	Moderate
Assessing Concept Mapping Competence Using Item Expansion-Based Diagnostic Classification Analysis [38]	Quasi-Experimental	Assessment Tool	Individual	Paper-Based	4/5	Moderate
The Effectiveness of Collaborative Mind Maps in Learning and Teaching Applied Technologies to Mathematics [63]	Mixed Methods	Learning Tool	Collaborative	Digital	5/5	Low
Determining the Effect of Concept Cards on Students' Perception of Physics Concepts with Concept Mapping [59]	Mixed Methods	Assessment Tool	Individual	Paper-Based	5/5	Low
Evaluating STEM-Based Sustainability Understanding: A Cognitive Mapping Approach [18]	Mixed Methods	Assessment Tool	Individual	Paper-Based	5/5	Low

Table A1. *Cont.*

Title	Research Design	Primary Applications	Learning Context	Predominant Format	Score (Yes/Total)	Quality (Risk of Bias)
From Science Class to Studio: Supporting Transformative Sustainability Learning Among Future Designers [33]	Mixed Methods	Assessment Tool	Mixed	Paper-Based	5/5	Low
Introducing Mindset Streams to Investigate Stances Towards STEM in High School Students and Experts [45]	Quasi-Experimental	Learning Tool	Individual	Both	5/5	Low

References

1. Vuorio, E.; Pernaa, J.; Aksela, M. Lessons for Sustainable Science Education: A Study on Chemists' Use of Systems Thinking Across Ecological, Economic, and Social Domains. *Educ. Sci.* **2024**, *14*, 741. [CrossRef]
2. Hamnell-Pamment, Y. The Role of Scientific Language Use and Achievement Level in Student Sensemaking. *Int. J. Sci. Math. Educ.* **2024**, *22*, 737–763. [CrossRef]
3. Sagarika, R.H. Syedkhamruddin Experiencing the Essence of Learning Database Management System Course Using C-Map Tool. *J. Eng. Educ. Transform.* **2020**, *33*, 460–464. [CrossRef]
4. Chen, C.H.; Chung, H.Y. Fostering Computational Thinking and Problem-Solving in Programming: Integrating Concept Maps Into Robot Block-Based Programming. *J. Educ. Comput. Res.* **2024**, *62*, 406–427. [CrossRef]
5. Buzan, T.; Buzan, B. *The Mind Map Book*; Pearson Education: London, UK, 2006; ISBN 1406612790.
6. Menouer, B.; Faouzi, L.; Ou Zennou, F.; Sbaa, M.; Nachit, B. Effects of Employing Mind Mapping Techniques in Geometry Instruction on Logical Thinking Abilities. *Int. J. Tech. Phys. Probl. Eng.* **2024**, *16*, 220–226.
7. Drigas, A.; Karyotaki, M. Learning Tools and Applications for Cognitive Improvement. *Int. J. Eng. Pedagog.* **2014**, *4*, 71–77. [CrossRef]
8. Waghmare, V.; Patil, S.; Mahajan, R.; Goudar, M. The Significance of Technology-Enhanced Learning towards Enhancing Engineering Education. *J. Eng. Educ. Transform.* **2024**, *37*, 713–718. [CrossRef]
9. Krab-Hüsken, L.E.; Pei, L.; de Vries, P.G.; Lindhoud, S.; Paulusse, J.M.J.; Jonkheijm, P.; Wong, A.S.Y. Conceptual Modeling Enables Systems Thinking in Sustainable Chemistry and Chemical Engineering. *J. Chem. Educ.* **2023**, *100*, 4577–4584. [CrossRef]
10. Jeong, I.; Evans, T. Knowledge Organisers for Learning: Examples, Non-Examples and Concept Maps in University Mathematics. *STEM Educ.* **2023**, *3*, 103–129. [CrossRef]
11. Susithra, N.; Deepa, M.; Reba, P.; Santhanamari, G. Coalescing Mind Maps as a Learning Aid Cum Formative Assessment Tool for Effective Teaching and Learning of Computer Architecture and Organization Course. *J. Eng. Educ. Transform.* **2023**, *36*, 236–243. [CrossRef]
12. Evans, T.; Jeong, I. Concept Maps as Assessment for Learning in University Mathematics. *Educ. Stud. Math.* **2023**, *113*, 475–498. [CrossRef]
13. Astriani, D.; Susilo, H.; Suwono, H.; Lukiati, B.; Purnomo, A.R. Mind Mapping in Learning Models: A Tool to Improve Student Metacognitive Skills. *Int. J. Emerg. Technol. Learn.* **2020**, *15*, 4–17. [CrossRef]
14. Fang, J.W.; He, L.Y.; Hwang, G.J.; Zhu, X.W.; Bian, C.N.; Fu, Q.K. A Concept Mapping-Based Self-Regulated Learning Approach to Promoting Students' Learning Achievement and Self-Regulation in STEM Activities. *Interact. Learn. Environ.* **2023**, *31*, 7159–7181. [CrossRef]
15. Drigas, A.; Papanastasiou, G.; Skianis, C. The School of the Future: The Role of Digital Technologies, Metacognition and Emotional Intelligence. *Int. J. Emerg. Technol. Learn. Online* **2023**, *18*, 65–85. [CrossRef]
16. Hwang, G.J.; Chang, S.C.; Song, Y.; Hsieh, M.C. Powering up Flipped Learning: An Online Learning Environment with a Concept Map-Guided Problem-Posing Strategy. *J. Comput. Assist. Learn.* **2021**, *37*, 429–445. [CrossRef]
17. Bredeweg, B.; Kragten, M.; Holt, J.; Kruit, P.; van Eijck, T.; Pijls, M.; Bouwer, A.; Sprinkhuizen, M.; Jaspar, E.; de Boer, M. Learning with Interactive Knowledge Representations. *Appl. Sci.* **2023**, *13*, 5256. [CrossRef]
18. Petrun Sayers, E.L.; Craig, C.A.; Skonicki, E.; Gahlon, G.; Gilbertz, S.; Feng, S. Evaluating STEM-Based Sustainability Understanding: A Cognitive Mapping Approach. *Sustainability* **2021**, *13*, 8074. [CrossRef]
19. Eppler, M.J. A Comparison between Concept Maps, Mind Maps, Conceptual Diagrams, and Visual Metaphors as Complementary Tools for Knowledge Construction and Sharing. *Inf. Vis.* **2006**, *5*, 202–210. [CrossRef]
20. Davies, M. Concept Mapping, Mind Mapping and Argument Mapping: What Are the Differences and Do They Matter? *High. Educ.* **2011**, *62*, 279–301. [CrossRef]

21. Pangandaman, H.K.; Datumanong, N.T.; Mukattil, N.P.; Hayudini, M.A.A.; Abdulhan, M.S.; Jilah, A.J.; Elam, K.S.; Abdurasul, J.N.A.; Najar, A.A.; Saradi, M.A. Effectiveness of Mind Mapping in the Improvement of Students Academic Performance: A Systematic Review. *Cuest. De Fisioter.* **2024**, *53*, 1363–1375.
22. Debbag, M.; Cukurbasi, B.; Fidan, M. Use of Digital Mind Maps in Technology Education: A Pilot Study with Pre-Service Science Teachers. *Inform. Educ.* **2021**, *20*, 47–68. [CrossRef]
23. Sebit, S.; Yıldız, S. Individual and Collaborative Computerized Mind Mapping as a Pre-Writing Strategy: Effects on EFL Students' Writing. *J. Comput. Educ. Res.* **2020**, *8*, 428–452. [CrossRef]
24. Kennedy, M.M. How Does Professional Development Improve Teaching? *Rev. Educ. Res.* **2016**, *86*, 945–980. [CrossRef]
25. Mongeon, P.; Paul-Hus, A. The Journal Coverage of Web of Science and Scopus: A Comparative Analysis. *Scientometrics* **2016**, *106*, 213–228. [CrossRef]
26. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*, 71. [CrossRef]
27. Hong, Q.N.; Pluye, P.; Fàbregues, S.; Bartlett, G.; Boardman, F.; Cargo, M.; Dagenais, P.; Gagnon, M.-P.; Griffiths, F.; Nicolau, B. Mixed Methods Appraisal Tool (MMAT), Version 2018. *Educ. Inf.* **2018**, *34*, 285–291.
28. Pace, R.; Pluye, P.; Bartlett, G.; Macaulay, A.C.; Salsberg, J.; Jagosh, J.; Seller, R. Testing the Reliability and Efficiency of the Pilot Mixed Methods Appraisal Tool (MMAT) for Systematic Mixed Studies Review. *Int. J. Nurs. Stud.* **2012**, *49*, 47–53. [CrossRef] [PubMed]
29. Veiga, F.; Gil-Del-Val, A.; Iriondo, E.; Eslava, U. Validation of the Use of Concept Maps as an Evaluation Tool for the Teaching and Learning of Mechanical and Industrial Engineering. *Int. J. Technol. Des. Educ.* **2024**, *35*, 383–401. [CrossRef]
30. Foley, R.W.; Ferguson, S.M.; Pollack, C.C. Measuring the Amorphous: Substantive and Methodological Outcomes from Concept Maps. *J. Eng. Educ.* **2021**, *110*, 161–183. [CrossRef]
31. Joshi, R.; Hadley, D.; Nuthikattu, S.; Fok, S.; Goldbloom-Helzner, L.; Curtis, M. Concept Mapping as a Metacognition Tool in a Problem-Solving-Based BME Course During In-Person and Online Instruction. *Biomed. Eng. Educ.* **2022**, *2*, 281–303. [CrossRef]
32. Dmoshinskaia, N.; Gijlers, H.; de Jong, T. Does Learning from Giving Feedback Depend on the Product Being Reviewed: Concept Maps or Answers to Test Questions? *J. Sci. Educ. Technol.* **2022**, *31*, 166–176. [CrossRef]
33. Bales, A.D.; Jensen, C.X.J.; Sekor, M.R.; Adinolfi, B. From Science Class to Studio: Supporting Transformative Sustainability Learning Among Future Designers. *Int. J. Sustain. High. Educ.* 2024; *ahead of print*. [CrossRef]
34. Dewi, H.I.; Hayun, M.; Susanto, A.; Zulfitria, Z. Analysis of the Effectiveness of Architectural Creative Learning Methods. *TEM J.* **2021**, *10*, 1190–1194. [CrossRef]
35. Fung, D.; Liang, T. The Effectiveness of Collaborative Mind Mapping in Hong Kong Primary Science Classrooms. *Int. J. Sci. Math. Educ.* **2023**, *21*, 899–922. [CrossRef]
36. Kontostavlou, E.Z.; Driga, A.M. Digital Technologies for Gifted Students' Education. *Glob. J. Eng. Technol. Adv.* **2023**, *15*, 191–204. [CrossRef]
37. Huang, H.M.; Tai, W.S.; Huang, T.C.; Lo, C.Y. Optimizing Inquiry-Based Science Education: Verifying the Learning Effectiveness of Augmented Reality and Concept Mapping in Elementary School. *Univers. Access. Inf. Soc.* **2024**, *24*, 681–694. [CrossRef]
38. Xia, S.; Zhan, P.; Chan, K.K.H.; Wang, L. Assessing Concept Mapping Competence Using Item Expansion-Based Diagnostic Classification Analysis. *J. Res. Sci. Teach.* **2024**, *61*, 1516–1542. [CrossRef]
39. Mohammed, R.; Kennedy-Clark, S.; Reimann, P. *Using Immersive and Modelling Environments to Build Scientific Capacity in Primary Preservice Teacher Education*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 6, ISBN 0123456789.
40. Thurn, C.M.; Hänger, B.; Kokkonen, T. Concept Mapping in Magnetism and Electrostatics: Core Concepts and Development over Time. *Educ. Sci.* **2020**, *10*, 129. [CrossRef]
41. Chen, K.F.; Hwang, G.J.; Chen, M.R.A. Knowledge Check-Based Concept Mapping in Digital Games: Impacts on Students' Learning Performance and Behaviors. *Educ. Technol. Res. Dev.* **2024**, *72*, 2297–2324. [CrossRef]
42. Alt, D.; Naamati-Schneider, L. Online Argumentation-Based Learning Aided by Digital Concept Mapping During COVID-19: Implications for Health Management Teaching and Learning. *Health Educ.* **2022**, *122*, 18–36. [CrossRef]
43. Reddy, S.N.; Pathlavath, M.; Narsareddygari, S.; Naik, S.M. Investigating The Transformative Effects of Active Learning Methodologies in The Field of Engineering Education to Improve Learning Outcomes in Students by Unleashing Their Potential. *J. Eng. Educ. Transform.* **2024**, *37*, 562–567. [CrossRef]
44. Drigas, A.; Mitsea, E.; Skianis, C. Metamemory: Metacognitive Strategies for Improved Memory Operations and the Role of VR and Mobiles. *Behav. Sci.* **2022**, *12*, 450. [CrossRef] [PubMed]
45. Brian, K.; Stella, M. Introducing Mindset Streams to Investigate Stances Towards STEM in High School Students and Experts. *Phys. A Stat. Mech. Its Appl.* **2023**, *626*, 129074. [CrossRef]
46. Vasanth, K.; Ravi, C.N.; Padmaja, A.; Prasad, M.R. A Novel Methodology for Improving Teaching Learning Process and Its Outcome on 2K Students for Engineering Education. *J. Eng. Educ. Transform.* **2020**, *33*, 323–328. [CrossRef]

47. Jackson, E.M.; Hanline, M.F. Using a Concept Map with RECALL to Increase the Comprehension of Science Texts for Children with Autism. *Focus Autism Other Dev. Disabil.* **2020**, *35*, 90–100. [CrossRef]
48. Rugh, M.S.H.; Capraro, M.M.; Capraro, R.M. Improving Self-Efficacy with Automatically Generated Interactive Concept Maps: DIME Maps. *Electron. J. E Learn.* **2023**, *21*, 141–157. [CrossRef]
49. Nitchot, A.; Gilbert, L.; Wettayaprasit, W. Pedagogically-Informed Knowledge Mapping: Representing Contextualised Competences and Technology Implemented. *Electron. J. E Learn.* **2021**, *19*, 308–320. [CrossRef]
50. Hwang, G.J.; Chen, Y.T.; Chien, S.Y. A Concept Map-Based Community of Inquiry Framework for Virtual Learning Contexts to Enhance Students' Earth Science Learning Achievement and Reflection Tendency. *Educ. Inf. Technol.* **2024**, *29*, 15147–15172. [CrossRef]
51. Hu, M.; Shealy, T.; Grohs, J.; Panneton, R. Empirical Evidence That Concept Mapping Reduces Neurocognitive Effort During Concept Generation for Sustainability. *J. Clean. Prod.* **2019**, *238*, 117815. [CrossRef]
52. Shah, S.M.; Elliott, C.; Nedungadi, P. Square Pegs and Round Holes: Pedagogy for Autistic Students in Computing Education. *IEEE Trans. Educ.* **2024**, *67*, 919–930. [CrossRef]
53. Liang, H.Y.; Hsu, T.Y.; Hwang, G.J. Promoting Children's Inquiry Performances in Alternate Reality Games: A Mobile Concept Mapping-Based Questioning Approach. *Br. J. Educ. Technol.* **2021**, *52*, 2000–2019. [CrossRef]
54. Hauck, D.J.; Melle, I. Molecular Orbital Theory—Teaching a Difficult Chemistry Topic Using a Cscl Approach in a First-Year University Course. *Educ. Sci.* **2021**, *11*, 485. [CrossRef]
55. Ma, J.; Liu, Q.; Yu, S.; Liu, J.; Li, X.; Wang, C. What Factors Influence Scientific Concept Learning? A Study Based on the Fuzzy-Set Qualitative Comparative Analysis. *Br. J. Educ. Technol.* **2024**, *56*, 250–275. [CrossRef]
56. Syarifuddin, H.; Atweh, B. The Use of Activity, Classroom Discussion, and Exercise (ACE) Teaching Cycle for Improving Students' Engagement in Learning Elementary Linear Algebra. *Eur. J. Sci. Math. Educ.* **2022**, *10*, 104–138. [CrossRef]
57. Goodwin, E.C.; Shapiro, C.; Freise, A.C.; Toven-Lindsey, B.; Moberg Parker, J. Synthesizing Research Narratives to Reveal the Big Picture: A CREATE(S) Intervention Modified for Journal Club Improves Undergraduate Science Literacy. *J. Microbiol. Biol. Educ.* **2023**, *24*, e00055-23. [CrossRef]
58. Becker, L.B.; Welter, V.D.E.; Aschermann, E.; Großschedl, J. Comprehension-Oriented Learning of Cell Biology: Do Different Training Conditions Affect Students' Learning Success Differentially? *Educ. Sci.* **2021**, *11*, 438. [CrossRef]
59. Kurt, H.S.; Sari, M. Determining the Effect of Concept Cards on Students' Perception of Physics Concepts with Concept Mapping. *Rev. Educ.* **2023**, *47*, 1–18. [CrossRef]
60. Sun, M.; Wang, M.; Wegerif, R.; Peng, J. How Do Students Generate Ideas Together in Scientific Creativity Tasks Through Computer-Based Mind Mapping? *Comput. Educ.* **2022**, *176*, 104359. [CrossRef]
61. Araya, R.; Collanqui, P. Are Cross-Border Classes Feasible for Students to Collaborate in the Analysis of Energy Efficiency Strategies for Socioeconomic Development While Keeping CO_2 Concentration Controlled? *Sustainability* **2021**, *13*, 1584. [CrossRef]
62. Simons, J.R. *Analyzing Educators' Conceptions of Stem Education: Using Concept Maps to Analyze Educators' Conceptions of STEM Education*; George Mason University: Fairfax, VA, USA, 2018.
63. Blass, L.; Rhoden, A.C. The Effectiveness of Collaborative Mind Maps in Learning and Teaching Applied Technologies to Mathematics. *Educ. Formação* **2024**, 1–22.
64. Kakoura, E.; Drigas, A. Digital Tools for Children with Reading Difficulties. *World J. Biol. Pharm. Health Sci.* **2023**, *14*, 129–136. [CrossRef]
65. Sanchiz, M.; Amadieu, F.; Lemarié, J.; Tricot, A. *Do Graphic and Textual Interactive Content Organizers Have the Same Impact on Hypertext Processing and Learning Outcome?* Springer: New York, NY, USA, 2023; Volume 35, ISBN 0123456789.

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

Computers Editorial Office
E-mail: computers@mdpi.com
www.mdpi.com/journal/computers

Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editor. 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 Editor 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 978-3-7258-7225-1