

Article

Advanced Analytics in Social Media Data Mining as a Driver of Digital Transformation in Cultural Heritage Tourism: The Case of Lamphun, Thailand

Pirapong Wongsensee ¹, Pintusorn Onpium ^{2,*}, Chakkrapong Kuensaen ¹ and Nantawan Muangyai ²

¹ International College of Digital Innovation, Chiang Mai University, Chiang Mai 50200, Thailand; pirapong.w@cmu.ac.th (P.W.); chakkrapong.k@cmu.ac.th (C.K.)

² Faculty of Humanities, Chiang Mai University, Chiang Mai 50200, Thailand; nantawan.m@cmu.ac.th

* Correspondence: pintusorn.o@cmu.ac.th; Tel.: +66-(0)-5394-3297

Highlights

What are the main findings?

- Chinese tourist UGC related to Lamphun reveals a predominance of positive sentiment across all ICH categories, with sentiment score emerging as the strongest predictor of tourism intention.
- Hybrid culinary heritage—lifestyle cafe exploring content generates the highest sentiment distribution and within cluster tourism intention rate.

What are the implications of the main findings?

- UGC analytics combining sentiment analysis, rule-based classification, and thematic clustering provides a scalable framework for underrepresented cultural destinations.
- Digital campaigns targeting Chinese audiences should integrate cultural heritage narratives with contemporary lifestyle activities to maximize engagement and tourism-intention conversion.

Abstract

Social media platforms and user-generated content (UGC) have become central to how travelers discover and evaluate cultural destinations, yet lesser-known second-tier heritage sites remain substantially underrepresented in digital tourism research. This study investigates how Chinese tourists perceive and engage with the intangible cultural heritage (ICH) of Lamphun, Thailand, through UGC collected from three major Chinese social media platforms (WeChat, Douyin, and Rednote) spanning the period from 2019 to 2023. A total of 642 relevant posts were analyzed using a mixed-methods analytical framework comprising SnowNLP-based Chinese-language sentiment analysis, rule-based tourism intention classification, and TF-IDF-driven K-means thematic clustering. Results indicate an overall predominance of positive sentiment, with sentiment score emerging as the strongest predictor of tourism intention. Thematic clustering revealed three distinct experiential dimensions, with culinary heritage and contemporary local lifestyle and cafe exploration generating the highest sentiment distribution and within-cluster tourism intention rate. These findings demonstrate the analytical value of integrated UGC data mining for underrepresented ICH destinations and offer empirical insights to support data-driven destination marketing strategies and destination management organization (DMO) decision-making for the promotion of secondary cultural heritage destinations.



Academic Editor: Dimitrios Theocharis

Received: 10 May 2026

Revised: 19 June 2026

Accepted: 22 June 2026

Published: 25 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: intangible cultural heritage tourism; user-generated content; Chinese social media; data mining; sentiment analysis; tourism intention; thematic clustering; AI technologies; Lamphun

1. Introduction

The digital transformation of tourism has fundamentally reshaped how prospective travelers discover and engage with destinations. Social media platforms and user-generated content (UGC) serve as important sources of information, destination discovery, and cultural trends, facilitating electronic word-of-mouth (eWOM) at an unprecedented scale and speed by enabling users to share personal experiences, opinions, generational perspectives, and travel recommendations with audiences instantaneously (Litvin et al., 2008; Di Gangi & Wasko, 2016). These developments have restructured the tourism ecosystem beyond mere technological adoption, fundamentally changing how tourists search for information, make decisions, and share experiences through real-time information exchange, peer-to-peer communication, and the democratization of travel knowledge (Xiang & Gretzel, 2010; Neuhofer et al., 2015; Nikolskaya et al., 2021). The exponential advancement of mobile technologies and pervasive internet connectivity has enabled tourists to engage continuously with digital platforms throughout the entire journey, from inspiration and trip planning to on-site experience and post-trip sharing, with online reviews and UGC also playing a significant role at each stage by providing essential information in the dynamic travel process (D. Wang et al., 2014; Zeng & Gerritsen, 2014; Bathla et al., 2024).

The advancement of artificial intelligence (AI) and data mining technologies has further expanded the analytical potential of social media data in tourism, enabling more systematic identification of patterns, sentiment, and tourist preferences from large volumes of UGC (Gau et al., 2024; Boukhari et al., 2025). The volume, velocity, and variety of tourism-related UGC have created both opportunities and challenges for destination marketers in understanding tourist perceptions and preferences, which has consequently become a central concern for researchers (Mariani et al., 2016; M. S. Lin et al., 2021). For destination management organizations (DMOs), insights derived from UGC analysis can be applied to develop more precisely targeted marketing content and communication strategies, with recent advances in generative AI further expanding the potential for adaptive destination storytelling and visitor engagement (Chemin et al., 2025; Sujatmiko et al., 2025; Nicolau, 2025).

Cultural heritage tourism represents a significant and growing segment of the global tourism market, reflecting a broader shift from product culture to process culture at diverse destinations (Gnoth & Zins, 2011; Richards, 2018; Espeso-Molinero, 2022). Tourists increasingly seek authentic, immersive experiences that allow deep engagement with local communities and traditions rather than merely observing cultural artifacts (Mkono, 2012; Wasela, 2023). This trend is particularly evident in the context of intangible cultural heritage (ICH), which reflects a fundamental change in tourist motivations from passive consumption of cultural products to active participation in authentic cultural experiences and the co-creation of meaning with local traditions and customs (Zayachka, 2023; Sharma & Aggarwal, 2024).

Among global outbound tourism markets, Chinese tourists represent a major source of international visitors and operate within a digital environment characterized by high mobile internet usage and rapid adoption of digital and social media (Quer & Peng, 2022; Lushchik, 2023; China Internet Network Information Center, 2023). The new generation of Chinese tourists demonstrates a strong preference for quality travel experiences, with

tourism activities, communication channels, and media identified as key factors in travel prioritization (Lojo, 2020; Kaewyu et al., 2020). Despite the strong presence of Chinese tourists in digital spaces, lesser-known destinations often lack visibility in Chinese social media discourse (H. Li et al., 2023).

Lamphun, a culturally rich city situated in the upper north of Thailand near Chiang Mai, the principal tourism center of northern Thailand, represents such an under-represented cultural tourism case. The city holds several registered intangible cultural heritage elements, including traditional textiles, rituals, and festivals. While Lamphun holds considerable cultural capital with significant tourism potential, this city attracts only approximately 1 million tourists per year compared to around 10 million annual visitors to the neighboring city of Chiang Mai. Notably, only around 2500 Chinese tourists visited Lamphun between 2018 and 2023 (Ministry of Tourism and Sports, 2023), underscoring the destination's limited visibility among Chinese travelers.

Significant research gaps persist despite the increasing importance of UGC analytics in cultural tourism, especially in relation to intangible cultural heritage. Studies in cultural tourism tend to gravitate towards well-established destinations, potentially leaving lesser-known cultural destinations underexplored (Chhabra et al., 2003; Alivizatou, 2012). In addition, limited research has combined UGC sentiment analysis with other analytical methods to examine intangible cultural heritage tourism engagement specifically through the lens of Chinese tourists and Chinese digital platforms. This study addresses these gaps by analyzing Chinese-language UGC data related to Lamphun, applying an integrated multi-method framework including sentiment analysis, tourism intention analysis, and ICH tourism theme clustering. The specific objectives are to: (1) examine sentiment patterns in Chinese social media discourse of Lamphun; (2) assess tourism-intention signals embedded in social media posts; and (3) identify thematic clusters and engagement patterns for destination marketing insights to support cultural heritage promotion of underrepresented cultural destinations. This study contributes to transformational digital tourism research by applying integrated UGC analytics to a lesser-known ICH destination, providing empirical insights that can inform data-driven destination marketing strategies and repositioning ICH experiences of underrepresented cultural destinations for broader tourism audiences.

2. Literature Review

2.1. Intangible Cultural Heritage (ICH) Tourism in Digital Contexts

Intangible cultural heritage (ICH) encompasses the practices, representations, expressions, knowledge, and skills recognized and shared by individuals, groups, or communities, which are passed down across generations as part of a living cultural identity (UNESCO, 2003). These cultural attractions include arts, architecture, historical and cultural heritage, gastronomy, literature, music, and other creative activities, which encompass the customs, values, beliefs, and way of life of a society (World Tourism Organization, 2019; Angelini et al., 2020). The growth of ICH tourism reflects a rising interest in experiential and culturally meaningful travel. This form of tourism is characterized by the visitor's primary motivation to learn, discover, and engage with tangible and intangible cultural resources of a destination (Al-Ababneh & Masadeh, 2019). Culturally motivated tourists tend to seek authenticity, community interaction, and cultural immersion when visiting heritage destinations, reflecting the pivotal role of intangible cultural experiences in contemporary cultural tourism (Seyfi et al., 2019; Sanjaya & Aruan, 2021; Richa & Naithani, 2024).

Digital information and technologies are also rapidly transforming cultural heritage tourism by providing innovative tools and techniques to engage visitors, create immersive experiences, and promote more sustainable preservation, including augmented and virtual reality experiences, digital storytelling, and immersive human–computer interaction (Abdo,

2019; Galani & Vosinakis, 2024; Mo et al., 2024; Z. Xu et al., 2025). These technologies and big data analytics have expanded the possibilities for understanding visitor engagement and strengthening ICH destination communication and marketing (Qiu et al., 2022). Specifically, AI-driven behavioral and emotional analytics allow marketers to deeply decode tourist perceptions and preferences, while advanced personalization and generative tools optimize destination branding and create highly interactive cultural recommendations and experiences (Deng, 2025; X. Hou & Paidi, 2025; Shao & Li, 2025). The digitization of cultural heritage content and the proliferation of social media platforms have generated substantial volumes of data that enable more systematic understanding of tourist perceptions, preferences, and engagement patterns (C. Y. Lin & Li, 2024). The intersection of digital technologies and cultural heritage has created new opportunities for destination marketing, visitor engagement, and heritage preservation. Data-driven approaches provide heritage destinations with systematic identification of tourist sentiment, online discourse patterns, and engagement behavior, offering empirical foundations for heritage promotion and management decisions (Abdelhamed, 2023; Yang & Shen, 2024; Chen, 2025).

Nevertheless, digital transformation of ICH tourism should be considered not merely as technological adoption for promotion, preservation, or management, but also as a process that mediates the representation and interpretation of cultural meanings in digital environments. Digital representation of ICH may not fully capture the participatory and context-specific nature of cultural practices (Y. Hou et al., 2022; Y. Zhang, 2025). Tensions between cultural authenticity, accessibility, and user engagement have been identified as challenges in digital ICH tourism contexts (Y. Zhang & Lee, 2021; J. Li et al., 2024; Z. Xu et al., 2025). This highlights the importance of examining how tourists perceive, discuss, and engage with ICH destinations through online platforms.

2.2. User-Generated Content (UGC) and Chinese Social Media

Social media and user-generated content (UGC) exert influence through specific roles in various stages and components of tourism decision-making, acting as need generators and supporters that inspire users to include destinations in future travel plans and strengthen the desire to travel (Ngai et al., 2014; X. Liu et al., 2019). Social media has fundamentally altered destination image formation by enabling tourists to access diverse, real-time perspectives from other travelers rather than relying solely on official marketing communications (Govers et al., 2007; Stepchenkova & Morrison, 2008). UGC sourced from online social media may surpass traditional marketing communications in credibility and persuasiveness (Fotis et al., 2012; Leung et al., 2013).

The growth of tourism in China has been shaped by internet adoption, which provides an efficient means of communicating cultural and historical heritage to prospective travelers (J. X. Ma et al., 2003; S. N. Z. Ma & Mohame, 2023). The digital behavior of Chinese tourists is distinctive, operating within a unique digital ecosystem in which major global platforms are inaccessible and domestic social media platforms operate with considerable independence (Yuan et al., 2026). Several domestic platforms dominate Chinese digital tourism discourse. WeChat functions as an all-in-one super application combining instant messaging, social networking, and mobile payment services. In contrast to open social platforms, electronic word-of-mouth (eWOM) on WeChat is largely restricted to primary social circles, as architectural privacy features encourage users to confine recommendations to immediate personal acquaintances (Pang et al., 2021). Douyin, known internationally as TikTok, has captured significant market share by enabling users to connect through entertainment, lifestyle, and information sharing via short video content (Tang et al., 2021). Xiaohongshu, also known as Rednote, has similarly gathered considerable momentum through picture and short video content creation, sharing, and interaction (Z. Wang et al.,

2022). These platforms combine social networking, content sharing, e-commerce, and travel planning functionalities in ways that influence the information-seeking behaviors and decision-making processes of Chinese tourists (L. Li et al., 2015; Basnyat & Jiahui, 2020).

In the context of cultural heritage tourism, studies of Chinese tourist UGC at heritage sites consistently reveal that content emphasizes emotional responses, cultural symbols, architectural features, and service quality, reflecting a combination of affective and cognitive engagement with heritage destinations (W. Liu et al., 2025; Qiu & Zhang, 2021). Emotionally expressive and authenticity-oriented UGCs more strongly influence destination image, visit intention, and heritage preservation awareness than purely factual information (H. Xu et al., 2021; Hua et al., 2024). Platform-specific patterns have also been identified, with Chinese tourists tending to post detailed reviews on travel platforms and emotionally expressive content on social media platforms such as Weibo and Xiaohongshu (Q. Yan et al., 2022; Y. Huang, 2024). Existing studies indicate that Chinese social media platforms can enhance the visibility of cultural destinations and facilitate the circulation of tourism experiences, particularly for smaller or lesser-known destinations (He, 2023). However, platform-based tourism practices may emphasize aesthetic visibility, short-term popularity, and repeated forms of online engagement, which may limit deeper cultural interpretation of heritage destinations (Tan, 2026). Understanding the specific characteristics and behavioral patterns of Chinese tourists across these platforms is therefore essential for examining how ICH destinations are perceived, represented, and discussed in Chinese digital tourism spaces.

2.3. Data Mining, UGC Analysis, and AI Technologies

As the use of social media platforms has increased dramatically, especially in the tourism industry, social media has become a valuable resource for analyzing social trends and opinions (Islam, 2021; Leelawat et al., 2022). In particular, younger-generation tourists largely depend on digital information to support decision-making (T. Wang & Park, 2023). Generation Z primarily interacts with brands through online platforms, with branding dimensions such as brand image, trust, and online brand experience significantly shaping their purchase intentions and highlighting the broader importance of digital engagement strategies in influencing this demographic's decision-making (Theocharis et al., 2025).

Big data mining has become an important capability in tourism marketing, as large datasets can be analyzed to predict trends and determine which marketing strategies are effectively targeted to different segments of consumer markets (Sawant & Desai, 2015). Artificial intelligence (AI) and machine learning further extend data mining approaches by supporting pattern recognition, classification, clustering, and prediction in large, complex, and unstructured datasets (Ge et al., 2017; Shu & Ye, 2022). The advancement of AI and data mining technologies has further expanded the analytical potential of social media big data in tourism, enabling more systematic identification of tourist behavioral patterns, preferences, and engagement trends (Nguyen et al., 2019; Rahmani et al., 2021).

Sentiment analysis has been used to analyze large sets of UGC, including online comments, social media posts, and consumer opinions, providing information regarding tourists' opinions and attitudes and yielding deeper insights into tourist experience preferences and consumer perceptions (B. Liu, 2012; Balahur et al., 2013; Borrajo-Millán et al., 2021). The analyzed sentiments shape the cognitive and affective dimensions of a destination's image, which directly govern tourist decision-making. Consequently, positive and emotionally engaging sentiment translates directly into increased intentions to visit or recommend a location (Bigne et al., 2019; Asyraff et al., 2023). Complex analytical approaches are required to capture the complexity, temporality, and heterogeneity of tourist perceptions, given the dynamic and co-created nature of destination image in social media environments (Choi et al., 2007; Marine-Roig & Clavé, 2015).

The application of such analytical approaches to social media data offers DMOs practical means of understanding tourist perceptions and informing evidence-based marketing decisions (Solazzo et al., 2021; A. Huang et al., 2022). Nevertheless, the effectiveness of these approaches, including AI-driven analysis, depends heavily on the volume and quality of available data, raising questions about the applicability in contexts where content is limited or fragmented (Q. Li et al., 2019; J. Liu et al., 2023).

2.4. Lamphun as a Cultural Tourism Case Context

Lamphun is situated in the upper northern region of Thailand, approximately 25 km south of Chiang Mai, the principal international tourism hub of northern Thailand (Figure 1). The city is characterized by a distinctive cultural landscape rooted in Lanna heritage, with major attractions including Haripunchai temple, Cham Thewi temple, and other historical and religious sites. Currently, Lamphun holds seven provincial-level and two national-level registered ICH elements, reflecting the continued presence of cultural practices. While some ICH elements remain relatively less recognized beyond local and domestic visitors, certain cultural elements have gained broader recognition and attract significant participation at both local and national levels through major festivals and events, including the Loy Krathong celebration, known as Hundred Thousand Lantern Festival, and Lamphun longan related tourism activities (Department of Cultural Promotion in Thailand, n.d.; Tourism Office, Lamphun City Hall, n.d.).

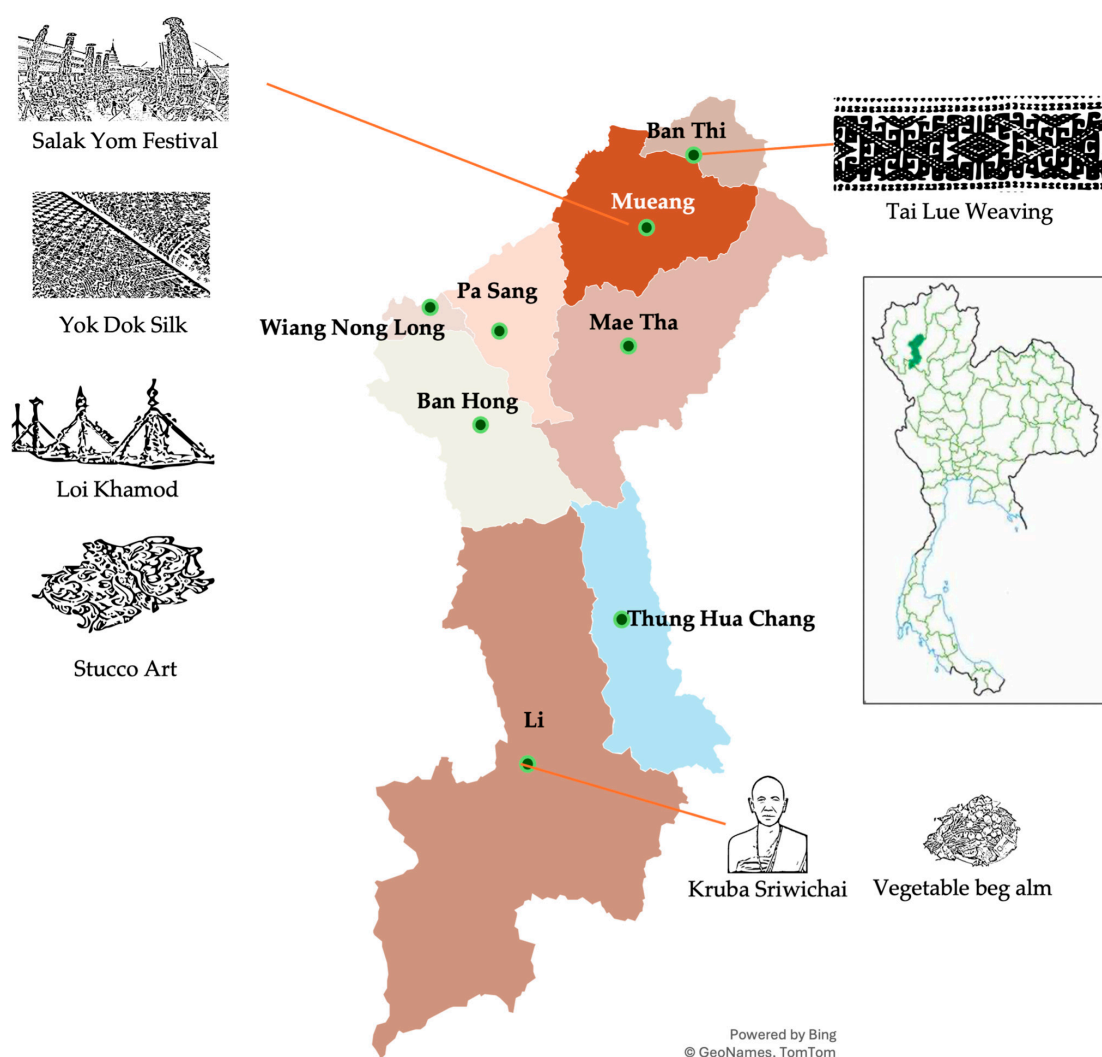


Figure 1. Location of Lamphun with selected registered intangible cultural heritage elements.

Secondary and rural tourism destinations compete through a combination of infrastructure quality, accommodation, community support, and a range of cultural activities and events (Lo et al., 2019). Destination competitiveness is further shaped by community ambiance, cultural values, and natural assets that influence tourist attitudes and behavior toward the destination (Tu et al., 2022). Despite this cultural richness, data-driven approaches to capturing tourist opinions and engagement with Lamphun remain limited, particularly from the perspective of Chinese digital platforms.

Collectively, the preceding review highlights several interconnected gaps in the existing literature. While digital transformation has expanded the possibilities for ICH destination marketing and visitor engagement, the application of data-driven analytics to lesser-known heritage destinations remains limited. UGC produced by Chinese tourists represents a distinctive and analytically valuable data source, yet existing studies have predominantly focused on well-known destinations, leaving secondary ICH destinations such as Lamphun relatively unexplored in Chinese digital tourism discourse. Furthermore, no prior study has applied an integrated multi-method framework specifically to examine how Chinese tourists perceive and engage with a lesser-known ICH destination through Chinese digital platforms. This study addresses these gaps by analyzing Chinese-language UGC related to Lamphun, offering empirical insights into how underrepresented cultural heritage destinations are perceived, represented, and discussed in digital tourism spaces.

3. Materials and Methods

3.1. Research Design and Overview

This research employs an advanced mixed-methods social media content analytics approach, collecting data from Chinese social media platforms, focusing on UGC related to Lamphun cultural heritage tourism. Web scraping techniques were employed to collect posts and comments by Chinese users from 2019 to 2023, yielding an initial dataset from three primary platforms: WeChat, Douyin (TikTok), and Rednote. The analytic framework integrates sentiment analysis with machine learning-based tourism intention analysis and thematic clustering. This multi-method approach aligns with contemporary best practices in tourism analytics, which emphasize integrating complementary techniques to provide comprehensive insights into complex phenomena (Mirzaalian & Halpenny, 2019; Gour et al., 2021). The selection of WeChat, Douyin, and Rednote reflects the dominant platforms used by Chinese users for information sharing, with each platform serving distinct functions (Yuan et al., 2026; Dai, 2023). The study follows a data mining and analytics framework (Figure 2) using Python-based tools (version 3.14.5) and the Orange Data Mining software (version 3.36.1) to systematically analyze UGC from WeChat, Douyin, and Rednote.

3.2. Data Collection

Social media data were collected through web-scraping techniques from WeChat, Douyin, and Rednote, focusing on posts and comments published between 2019 and 2023. Keywords used in the scraping included terms in the Chinese language equivalent to 'Lamphun tourism,' 'Cultural heritage in Lamphun,' and specific cultural elements, for example, 'Lamphun Lantern Festival.' Data were structured as spreadsheets containing timestamped entries, pseudonymized usernames, content text, engagement metrics (likes, comments, favorites, and shares), and topic labels. All original content was in Chinese.

3.3. Data Cleansing

The collected Chinese-language content underwent data preprocessing, involving the removal of structural noise, elimination of duplicate entries, text filtering, and exclusion of

incomplete records. The cleaned data were subsequently structured into CSV format for further analytical procedures.

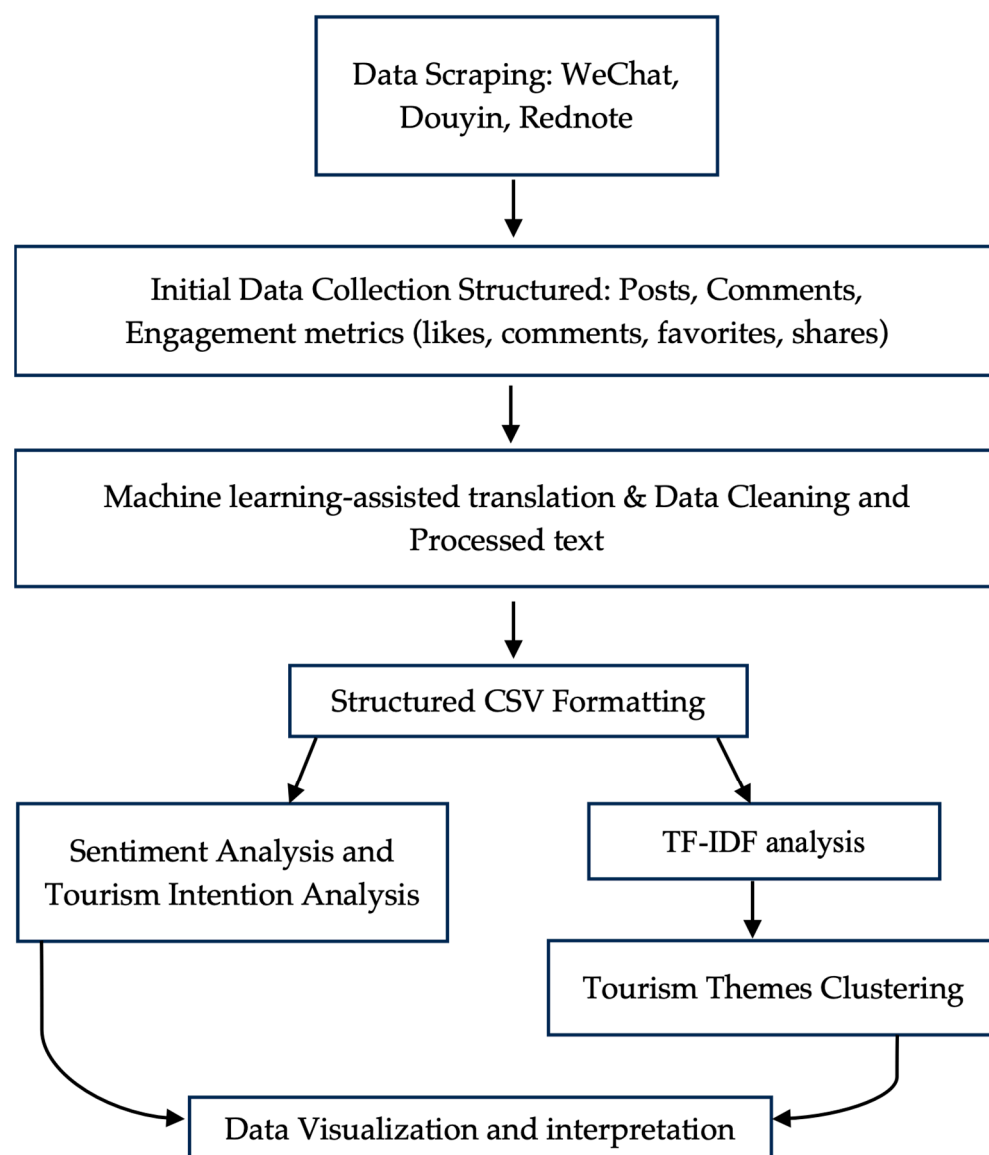


Figure 2. Overview of the data mining and analytical framework employed in this study.

3.4. Data Analysis

3.4.1. Sentiment Analysis and Tourism Intention

Sentiment analysis was conducted using SnowNLP, a Python-based natural language processing library specifically developed for Chinese-language text analysis. Sentiment classification was performed directly on the original Chinese-language UGC, preserving semantic integrity and reducing potential distortion associated with translation (G. Yan et al., 2014). SnowNLP produces a probability-based sentiment score ranging from 0 to 1. Based on this scoring range, sentiment was classified into three categories: positive (score ≥ 0.60), neutral (score 0.40–0.59), and negative (score < 0.40). Sentiment scores were computed and classified across the five UNESCO ICH categories relevant to Lamphun’s cultural tourism.

To investigate the relationships among UGC sentiment, engagement metrics, and the expression of tourism intention, a semantic analysis of UGCs was conducted. Tourism intention signals were operationalized as a binary variable, determined through a dictionary-

assisted rule-based classification applied to the dataset (Abel & Lantow, 2019; Hamroun & Gouider, 2020; Y. Liu & Beldona, 2021). Posts were classified as intention-driven based on the presence of explicit pre-defined indicators of tourism motivation, such as expressions of willingness to visit, recommendations, references to festival attendance, or evidence of travel planning. Posts that described past experiences, information, or general destination attributes without explicit prospective travel intention were classified as non-intentional.

The tourism-intention classifier keyword set was developed through a systematic iterative process. Initial terms were derived from prior literature related to tourism intention, followed by an exploratory review of the sub-corpus to identify domain-specific vocabulary. Destination-specific expressions relate to Lamphun, and itinerary planning was subsequently incorporated, and exclusion rules were applied to minimize false positives. The complete keyword list, source references, and iteration log are documented in Supplementary Materials. To validate the classifier labels, a stratified random sample comprising 30% of the corpus across platforms and ICH categories was independently coded by two annotators. Inter-coder reliability was assessed using Cohen's κ . Disagreements were resolved through adjudication by a third independent annotator (Dligach et al., 2010; Bayerl & Paul, 2011). The finalized labels constituted the gold-standard reference set. Classifier performance was evaluated against the gold-standard subset using a confusion matrix and standard binary classification metrics.

Classification was implemented programmatically and integrated into the cleaned dataset. A chi-square test of independence was performed to assess the relationship between sentiment and tourism intention, followed by pairwise comparisons using Fisher's Exact Test for specific group contrasts, with Bonferroni correction applied to adjust for multiple comparisons. To address platform heterogeneity, chi-square tests of independence were additionally conducted separately within each platform. Temporal stratification was conducted using three periods, namely the pre-pandemic restriction period (posts published before 26 March 2020), the pandemic restriction period (posts published from 26 March to 9 January 2023), and the post-pandemic re-opening period (posts published from 10 January 2023 onwards). This stratification was used to examine the stability of the sentiment-intention relationship across periods.

To evaluate the relationship between engagement metrics (likes, comments, favorites, shares) and intention, non-parametric Mann–Whitney U tests were applied. All metrics were standardized within each platform using z-scores before pooling, to account for structural differences across different platforms. Given the potential non-linearity and multicollinearity in the engagement data, a Random Forest classifier was used to explore the predictive capacity of sentiment and platform-normalized engagement features for tourism intention.

3.4.2. Term Frequency-Inverse Document Frequency (TF-IDF) Analysis and Tourism Theme Clustering

The Chinese-language cleaned textual content was translated into English using machine learning-assisted translation to support international interpretability of the thematic clustering. The translated text was subsequently used for TF-IDF analysis. The analysis was performed using the Orange Data Mining software. Thematic clustering was conducted using K-means applied to TF-IDF keyword vectors. K-means clustering is an unsupervised machine learning algorithm that partitions documents into clusters based on the similarity of their TF-IDF vectors, enabling the discovery of thematic structure in large text corpora without predefined categories (Jain, 2009; Kwale, 2013; Z. Li et al., 2019). K-means has proved effective for identifying tourism themes, destination attributes, and tourist experience dimensions from social media content (Arefieva et al., 2021; Harish et al., 2023). The optimal number of clusters was determined using the Elbow method and additional cluster

validation indices. Candidate solutions from $K = 2$ to $K = 8$ were compared using the silhouette score, Davies-Bouldin index, gap statistic, and bootstrap stability based on the adjusted Rand index. The final cluster was selected by considering both statistical validation results and thematic interpretability. The K-means algorithm was run with 100 initializations and a fixed random seed of 42 to ensure reproducibility. High-ranking TF-IDF terms from cluster centroids were used to characterize the thematic structure of each cluster. Cluster-level sentiment was summarized using the proportional distribution of sentiment categories within each cluster. Tourism intention was calculated as the proportion of intentional posts assigned to the same cluster.

3.4.3. Data Visualization and Interpretation

Frequently occurring terms within the dataset were visualized using a word cloud, highlighting core themes and tourist interests. Keyword prominence corresponded directly to frequency of mentions, enabling the identification of key themes and concepts relevant to tourism practitioners and destination managers. Additionally, this visualization assisted policymakers and tourism managers in decision-making by highlighting overarching themes that indicate areas of significant value or those requiring improvement (Filatova, 2016; Dobrea et al., 2023). Principal Component Analysis (PCA) was subsequently applied to the TF-IDF feature matrix to reduce dimensionality and visualize the K-means clusters in two dimensions, providing a visualization of data distributions and relative cluster positioning.

4. Results

A total of 5663 posts and more than 10,000 comments were collected during the 2019–2023 period. The data spanned multiple sources, user types, and temporal frames. After filtering for relevant keywords, the dataset was refined to 642 contents directly pertaining to the research focus.

4.1. Sentiment Analysis and Tourism Intention

4.1.1. Sentiment Analysis

The results revealed significant patterns in Chinese tourists’ perceptions of Lamphun, indicating an overall prevalence of positive sentiment, with 392 positive reviews across all components, yielding an average sentiment score of 0.6756. The analysis also identified 115 neutral reviews and 135 negative reviews within the dataset. Notably, different aspects of the tourism experience elicited varying sentiment distributions. ‘Knowledge and practices concerning nature and the universe’ garnered 200 positive reviews with the lowest average sentiment score of 0.6532, along with 70 neutral and 75 negative reviews, representing the highest proportion of negative sentiment among all categories. ‘Traditional craftsmanship’ achieved the highest average sentiment score (0.7886), recording 15 positive, two neutral, and two negative reviews. Overall, the sentiment of posts toward Lamphun was predominantly positive, as summarized in Table 1.

Table 1. Sentiment analysis of posts related to Lamphun, categorized by UNESCO’s ICH elements.

ICH Elements	Sentiment Classification	Review Volume	Average Sentiment Score
Oral traditions and expressions	Positive	42	0.7108
	Neutral	9	
	Negative	12	
Performing arts	Positive	32	0.7676
	Neutral	5	
	Negative	5	

Table 1. Cont.

ICH Elements	Sentiment Classification	Review Volume	Average Sentiment Score
Social practices, rituals, and festival events	Positive	103	0.6729
	Neutral	29	
	Negative	41	
Knowledge and practices concerning nature and the universe	Positive	200	0.6532
	Neutral	70	
	Negative	75	
Traditional craftsmanship	Positive	15	0.7886
	Neutral	2	
	Negative	2	
Overall Lamphun reviews	Positive	392	0.6756
	Neutral	115	
	Negative	135	

4.1.2. Validation of the Tourism-Intention Classifier

The tourism-intention classifier was evaluated using a gold-standard subset of 193 posts. Inter-coder reliability between the two independent annotators reached $\kappa = 0.7343$, indicating substantial agreement (Landis & Koch, 1977). Following adjudication, classifier performance against the finalized gold standard labels was summarized using a confusion matrix (Figure 3) and binary classification metrics (Table 2).

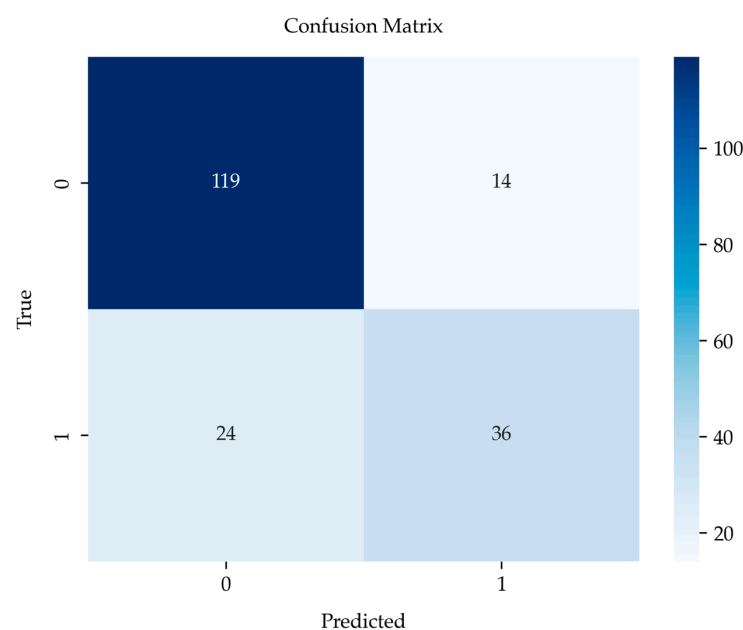


Figure 3. Confusion matrix comparing the rule-based tourism-intention classifier with the gold-standard labels. Note. 0 = non-intention; 1 = Intention.

Table 2. Binary classification performance of the tourism-intention classifier.

Metric	Value
Accuracy	0.803
Balanced accuracy	0.747
Precision	0.720
Recall/Sensitivity	0.600
Specificity	0.895
F_1 -score	0.655
Matthews correlation coefficient (MCC)	0.523

4.1.3. Sentiment and Tourism Intention Analysis

The cross-tabulation showed that tourism intention-driven posts were unevenly distributed across three categories (Table 3).

Table 3. Cross-tabulation of sentiment category and tourism intention.

Sentiment Categories	Intention-Driven Posts	Non-Intention Driven Posts	Intentional Rate (%)
Positive	74	318	18.88%
Neutral	12	123	8.89%
Negative	6	109	5.22%

A chi-square test of independence indicated a statistically significant association between sentiment and tourism intention ($\chi^2 = 17.6375$, $p < 0.001$). Pairwise Fisher's exact test revealed significant differences in tourism intention between positive and negative sentiment groups and between positive and neutral sentiment groups, whereas the differences between the neutral and negative groups were not statistically significant. The pairwise results are summarized in Table 4 and Figure 4.

Table 4. Fisher's exact test for pairwise comparisons of tourism intention by sentiment category.

Sentiment Categories	Odds Ratio	<i>p</i> -Value	Bonferroni-Adjusted <i>p</i> -Value
Positive vs. Negative	2.385	0.0066	0.0198 *
Positive vs. Neutral	4.227	0.0002	0.0006 *
Neutral vs. Negative	1.772	0.3298	0.9894

* indicates statistically significant.

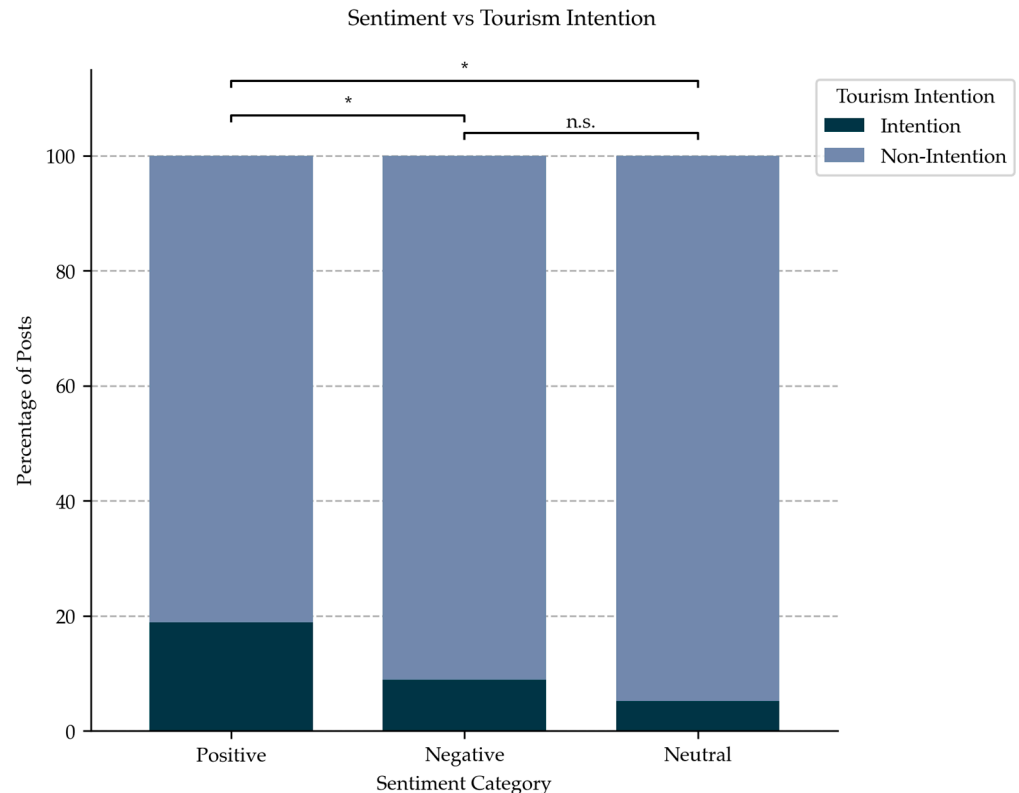


Figure 4. Distribution of tourism intention post across sentiment categories with Fisher's exact test results (* indicates statistically significant; Bonferroni-adjusted *p* value < 0.05, n.s. = non-significant).

4.1.4. Platform and Temporal Stratification of Sentiment-Intention Analysis

To account for structural differences across social media environments, the data were stratified across three platforms: WeChat, Douyin, and Rednote. The distribution of posts across platforms is presented in Table 5.

Table 5. Distribution of posts and tourism-intention classification by platform.

Platforms	Number of Post (%)	Intention-Driven Posts	Non-Intention Driven Posts
WeChat	463 (72.1%)	36	427
Douyin	64 (10%)	10	54
Rednote	115 (17.9%)	46	69

The platform-stratified analyses indicated that the sentiment-intention association was not consistent across platforms. No statistically significant association was observed on WeChat ($\chi^2 = 1.706$, $p = 0.426$). or Douyin ($\chi^2 = 0.836$, $p = 0.658$). In contrast, a statistically significant association was retained on Rednote ($\chi^2 = 9.523$, $p = 0.009$).

Temporal-stratified analysis revealed variation in the sentiment-intention association across periods. No statistically significant association was observed during the pre-pandemic restriction period ($\chi^2 = 2.972$, $p = 0.226$) or the post-pandemic reopening period ($\chi^2 = 3.407$, $p = 0.182$). A statistically significant association was observed during the pandemic restriction period ($\chi^2 = 10.981$, $p = 0.004$).

4.1.5. Engagement Metrics and Random Forest Classification

Platform-normalized engagement metrics showed statistically significant differences between intention-driven and non-intention posts for likes, favorites, and shares. No statistically significant difference was observed for Comments (Table 6).

Table 6. Mann–Whitney U tests results for platform-normalized engagement metrics by tourism intention classification.

Platform-Normalized Engagement Metric	U-Statistic	p-Value
Likes	29,037.5	0.019 *
Comments	26,785.0	0.254
Favorites	20,909.5	<0.001 *
Shares	22,589.0	0.037 *

* indicates statistically significant p -value < 0.05.

The Random Forest classifier achieved an overall classification accuracy of 83.9%, a balanced accuracy of 63.4%, a Matthews correlation coefficient (MCC) of 0.291, and a class-specific F_1 -score of 0.381 for the tourism-intention category, as detailed in Table 7.

Table 7. Random Forest classification performance by class.

Classification	Precision	Recall	F_1 -Score	Support
With intention	0.421	0.348	0.381	23
Without intention	0.894	0.920	0.907	138
Accuracy			0.839	161
Macro avg.	0.658	0.634	0.644	161
Weighted avg.	0.827	0.839	0.832	161

Random Forest feature importance scores indicated that sentiment contributed most substantially to the classification of tourism intention, with an importance score of 0.541, followed by platform-normalized engagement metrics and platform variables (Table 8).

Table 8. Random Forest feature importance score for tourism intention classification.

Feature	Importance
Sentiment score	0.541
Likes	0.136
Comments	0.083
Favorites	0.074
Shares	0.055
Rednote	0.052
WeChat	0.051
Douyin	0.008

4.2. TF-IDF and Tourism Theme Clustering

4.2.1. TF-IDF and Word Cloud Visualization

The TF-IDF analysis identified the most prominent terms in the translated corpus with ‘festival,’ ‘chiangmai,’ and ‘lanterns’ ranking highest by TF-IDF score (Table 9).

Table 9. Top 20 words with TF-IDF Score.

Words	TF-IDF Score	Words	TF-IDF Score
festival	0.037	beautiful	0.017
chiangmai	0.037	food	0.015
lanterns	0.033	travel	0.014
epidemic	0.029	night	0.013
day	0.024	cafe	0.013
delicious	0.022	restaurant	0.012
pot	0.020	meat	0.012
lanterns festival	0.020	longan	0.011
times	0.019	favorite	0.011
temple	0.018	photos	0.011

The word cloud visualization, shown in Figure 5, depicted frequently occurring keywords in Chinese tourist-generated content about Lamphun. The term ‘chiangmai’, referring to Chiang Mai, appeared prominently in the center of the word cloud. Other frequently mentioned terms included ‘festival,’ ‘lanterns,’ ‘temple,’ ‘travel,’ ‘cafe,’ and ‘city.’



Figure 5. Word cloud visualization of translated Lamphun-related user-generated content.

4.2.2. Tourism Themes Clustering and PCA Visualization

Validation indices across candidate solutions from $K = 2$ to $K = 8$ are presented in Table 10. Although no single K value was optimal across all indices, the three-cluster solution ($K = 3$) exhibited the highest bootstrap stability, with silhouette and gap statistic values comparable to the highest-performing candidate solutions. Accordingly, $K = 3$ was selected based on statistical validation outcomes and thematic interpretability.

Table 10. K-means cluster validation indices for candidate cluster solutions.

K	Silhouette Score	Davies-Bouldin Index	Gap Statistic	Bootstrap ARI
2	0.087	4.213	3.413	0.685
3	0.088	5.432	3.422	0.759
4	0.089	4.897	3.428	0.683
5	0.089	4.693	3.434	0.664
6	0.090	4.681	3.437	0.598
7	0.087	4.722	3.439	0.483
8	0.089	4.248	3.436	0.313

ARI = adjusted Rand index.

A total of three thematic clusters were identified, each representing a semantically coherent group of tourist-generated content, as summarized in Table 11. Cluster 1 accounted for the largest proportion of the corpus ($n = 483$) and was characterized by terms related to local food and local lifestyle experiences, with a positive sentiment distribution of 55.3% and the lowest within-cluster tourism intention rate across the three clusters (7.7%). Cluster 2 ($n = 91$) was associated with cafe culture, local cuisine, and religious-related sites, and recorded the highest positive sentiment distribution (81.3%) and the highest within-cluster tourism intention rate (45.1%). Cluster 3 ($n = 68$) was characterized by terms related to major festivals and ritual symbolism of Lamphun, with a positive sentiment distribution of 75.0% and a within-cluster tourism intention rate of 20.6%.

Table 11. Cluster profiles by theme, keywords, sentiment distribution, and within-cluster tourism intention rate.

Cluster	Suggested Theme	Keywords	Sentiment Distribution (%)	Within-Cluster Tourism Intention Rate (%)
1	Local culinary exploring and local lifestyle experiences	pot, times, epidemic, meat, lantern, photos, rice, pork, food, restaurant	Positive: 55.3% Neutral: 22.5% Negative: 22.2%	7.7%
2	Cafe culture, local cuisine, and religious-related sites	chiangmai, cafe, temple, travel, noodles, coffee, garden, longan, experience, day	Positive: 81.3% Neutral: 2.2% Negative: 16.5%	45.1%
3	Major festival and ritual symbolism	lanterns, festival, chiangmai, life, temple, sky, daily, night, thousand, beautiful	Positive: 75.0% Neutral: 5.9% Negative: 19.1%	20.6%

The PCA projection illustrated the spatial distribution of the three K-means thematic clusters in a two-dimensional component space (Figure 6). The projection visualized the relative positioning of posts assigned to each cluster, with partial overlap among thematic groups. Cluster 1 exhibited the widest distribution and partially overlapped with Cluster 2. Cluster 3 appeared more peripheral in the PCA projection, with greater separation from the remaining clusters.

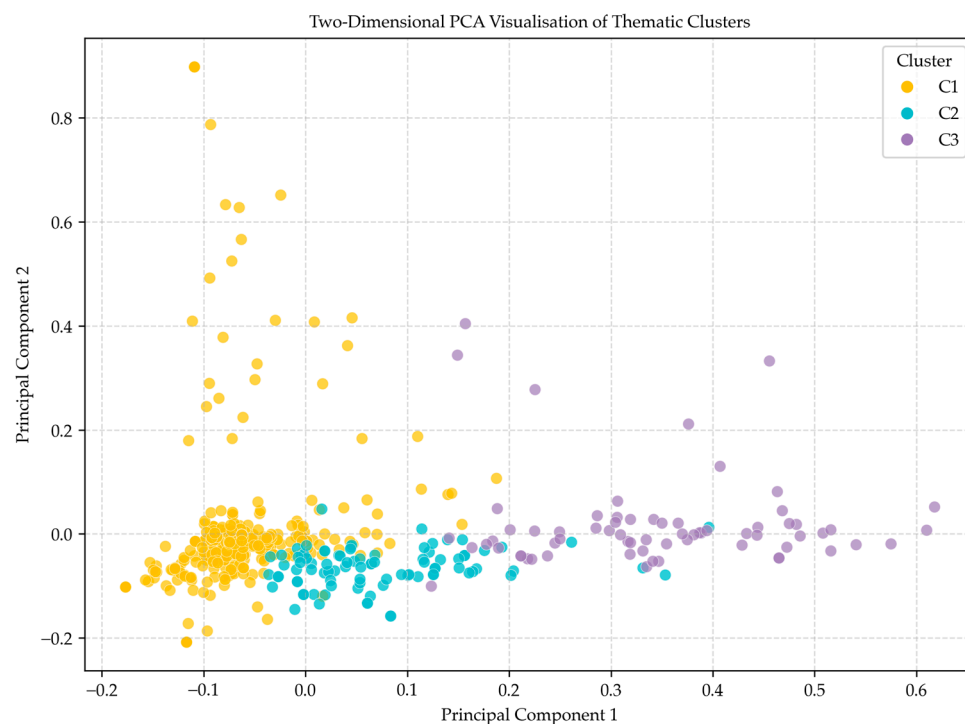


Figure 6. Two-dimensional PCA representation of TF-IDF Matrix for Thematic Clustering. Each point represents a tourism-related content entry distributed along the first and second principal components (PC1 and PC2), with colors indicating cluster membership as determined by K-means analysis.

5. Discussion

This study applies an integrated UGC analytics framework to examine destination perception and tourism intention among Chinese visitors to a second-tier intangible cultural heritage destination in northern Thailand, contributing empirical evidence in the field of digital heritage tourism research. The predominance of positive sentiment across the corpus indicates favorable overall perceptions of Lamphun as a cultural heritage destination among Chinese users. Among the UNESCO ICH categories, ‘Traditional craftsmanship’ recorded the highest average sentiment score, while oral traditions and performing arts also yielded comparatively high scores despite relatively limited content volume. Digital representation of cultural elements presents inherent communicative challenges in social media platforms (Y. Hou et al., 2022; Y. Zhang, 2025). The limited digital visibility of these cultural elements consequently suggests structural constraints of the platforms rather than a lack of tourist interest. Recent cultural tourism trends are shifted toward intangible cultural experience and meaningful cultural engagement, particularly when cultural expressions contain visual, symbolic, and interactive aspects (Richards, 2018; J. Li et al., 2024). This pattern suggests that targeted digital content or social media strategies could increase the visibility of positively received but underrepresented cultural elements among Chinese audiences (Wanju et al., 2025).

The categories of ‘Knowledge and practices concerning nature and the universe’ generated the largest amount of content, followed by ‘Social practices, rituals and festival events’. Both categories are associated with the culinary culture and festivals, highlighting their importance as vehicles for connecting tourists with the local community. This demographic has demonstrated a strong willingness to engage with local food culture and cultural events as part of the cultural tourism experience. This finding aligns with studies that emphasized the role of festivals and culinary experiences in cultural tourism engagement (Gnoth & Zins, 2011; López-Guzmán et al., 2016). The positive engagement with festivals and culinary experiences is a key driver of cultural tourism engagement,

leading to enhanced visitor satisfaction, loyalty, and a stronger destination image (Iswanto & Irsyad, 2024; Chendraningrum & Suryana, 2025). However, both categories recorded elevated proportions of negative sentiment. This finding may indicate that certain elements of Lamphun's cultural offerings do not fully align with Chinese tourist expectations. Differences in gastronomic habits and culinary traditions have been identified as genuine challenges for Chinese visitors (B. Zhang & Niyomsilp, 2020), and the availability and accessibility of varied local cuisine represent important factors in shaping destination image and overall visitor satisfaction among Chinese tourists (Satchapappichit, 2020; J. Cheng & Jiang, 2024). For DMOs, these findings suggest a concrete opportunity: targeted communication that contextualizes Lamphun's culinary heritage within frameworks familiar to Chinese visitors, including comparisons with regional flavors or emphasis on widely appreciated local dishes, could reduce perceptual distance and address negative sentiment drivers.

The analysis of tourism intention signals revealed a statistically significant association between positive sentiment and travel intention. The present findings support prior research indicating that a positive destination image strongly influences intention to visit or recommend (Afshardoost & Eshaghi, 2020), and favorable emotional evaluations strengthen attitudes and subsequently increase the likelihood of the intended action (Lam & Hsu, 2005; Kiatkawsin & Han, 2017). However, the platform-stratified analysis revealed that the sentiment-intention association was statistically significant on Rednote, but not on other platforms. This platform-specific pattern reflects the distinct communicative functions of each platform within the Chinese digital ecosystem. Rednote operates as a lifestyle discovery and travel inspiration platform where positive sentiment appears to be more directly associated with travel consideration. Platform's recommendations directly guide consumption behavior, and users frequently develop specific travel tendencies based on curated content (J. Liu, 2024; Zou et al., 2025). These findings provide quantitative evidence that Rednote represents a particularly high-value channel for DMOs seeking to convert positive digital engagement into travel intention among Chinese audiences interested in cultural heritage destinations. The temporal stratification revealed significant variations in sentiment-intention patterns across different time periods, with the pandemic restriction period showing particular statistical significance. This suggests that positive destination-related content may reflect deferred travel interest and future travel aspiration. Travel restrictions have been shown to produce a dual effect, suppressing immediate travel plans while simultaneously intensifying future travel desire, a pattern described as 'compensatory' or 'revenge' travel driven by the emotional and cognitive need to compensate for lost experiences (Kim et al., 2021; Y. Zhang et al., 2021; J. Wang & Xia, 2021). The continued digital engagement with Lamphun's cultural heritage during travel restrictions suggests that the destination retained a meaningful presence in digital aspirations of Chinese visitors, highlighting the role of digital platforms in maintaining destination awareness during the period of low actual visitation.

Platform-normalized likes, favorites, and shares showed significant differences between intention and non-intention posts, indicating that interactive engagement amplifies both the reach and persuasive value of user-generated content. Engagement behaviors also serve as social endorsement mechanisms, amplifying the reach and impact of UGC by encouraging further interaction and content dissemination, increasing the perceived credibility and appeal of the destination (Aljarah et al., 2022; Kaukuntla, 2025). High engagement signals authenticity and trustworthiness, making content more persuasive and more likely to influence travel attitudes and behaviors (Mathur et al., 2021; Konak, 2024), and interactive digital responses have been identified as mediators between exposure to tourism content and decision-making (Cheung et al., 2021). The Random Forest model indicated that sentiment score was the most important predictor of tourism intention,

followed by platform-normalized engagement metrics and platform variables. Predictive analytics studies have reported that models integrating affective and engagement variables demonstrated superior accuracy in forecasting travel intentions compared to sentiment-only approaches (Y. Li et al., 2022). The present results extend this finding to the specific context of Chinese-language UGC. However, engagement indicators should be interpreted with caution. Active engagement behavior has not been shown to significantly impact destination advocacy or visit intention (Tri et al., 2023), and subjective or unstructured variables such as comments tend to exhibit lower predictive consistency due to the heterogeneous content, including questions, clarifications, or criticisms that do not necessarily indicate travel intention (Hagenauer & Helbich, 2017; L. Cheng et al., 2018). Similarly, sharing behavior is driven by a complex and inconsistent set of motivational factors that limit predictive power (Tsekouropoulos, 2019). The model results suggest that machine learning approaches can be used to explore relationships among affective expressions, engagement features, and tourism-intention signals in social media data, offering an exploratory analytical complement to survey-based methods (M. S. Lin et al., 2021; N. Li et al., 2025). For Lamphun, these findings suggest that destination communication should prioritize positively framed cultural heritage content, while engagement metrics should be used as supplementary indicators for monitoring audience response and potential travel interest.

The TF-IDF and clustering results demonstrated that Lamphun's digital destination image is multidimensional, encompassing cultural festivals, religious heritage, culinary experiences, and contemporary lifestyle activities. The prominent co-occurrence of "chiang-mai" indicates that many Chinese visitors perceive Lamphun as part of a broader regional itinerary centered on Chiang Mai rather than as an independent destination. Tourism statistics confirm substantially higher international visitation rates to Chiang Mai (Ministry of Tourism and Sports, 2023), and this perceived association points toward the potential value of collaborative regional destination marketing strategies in which DMOs from Lamphun and Chiang Mai develop coordinated content campaigns that leverage the established appeal of Chiang Mai to drive secondary visitation to Lamphun. Tourists seeking alternative experiences beyond heavily visited centers may venture into second-tier cities owing to distinctive cultural attractions, local food, culture, and more affordable prices (Fakfare et al., 2020). However, the benefits of this spillover effect are not uniform; destinations with significant heritage assets tend to benefit more substantially from the overflow of tourism from larger urban centers compared to those with fewer major heritage elements (Tian et al., 2021). Lamphun, with its rich intangible cultural heritage, is well-positioned to capitalize on this dynamic.

The thematic clustering revealed three distinct clusters that reflect the interplay between cultural specificity, experiential diversity, and temporal context in shaping destination perception among Chinese tourists. Cluster 1, centered on culinary experiences and lifestyle exploring, accounted for the largest share of the corpus and recorded the lowest tourism-intention rate. This pattern suggests that everyday cultural content generated broad digital visibility, although such visibility did not necessarily translate into explicit travel intention. In contrast, Cluster 2, combining cafe exploring, local cuisine, and religious-related sites, recorded both the highest positive sentiment distribution and the highest tourism-intention rate. This reflects the growing interest of Chinese tourists in contemporary leisure experiences, suggesting that Lamphun's appeal extends beyond traditional heritage tourism into lifestyle-oriented activities. The outcomes align with the evolving preferences of younger Chinese tourists. Gao et al. (2021) found that younger Chinese tourists favored contemporary leisure activities that complement the historical significance of chosen destinations. This shift indicates a broader movement toward integrating cultural exploration with leisure pursuits. This trend also indicates a diversification

of travel motivations beyond traditional sightseeing (Hu & Chen, 2023). Cluster 3 was defined by festival-centered narratives, particularly centered on the festival, temple rituals, and lantern symbolism. These elements embody affective engagement and symbolic experiences, resonating with the notion that participation in culturally embedded rituals strengthens travel motivation. The study revealed that sensory immersion, narrative symbolism, and opportunities for cultural participation associated with the festival experience significantly enhance the memorability of heritage encounters, ultimately strengthening behavioral intention among visitors (S. Ye et al., 2025). Consistent with W. Wang et al. (2015), the findings of this study determined that Chinese tourists are driven by religious faith, desire for cultural enrichment, and need for mental rejuvenation when visiting religious heritage sites. Collectively, these cluster-level findings offer DMOs a prioritized content framework, festival and ritual-driven narratives should anchor digital campaigns targeting Chinese audiences, with hybrid heritage—lifestyle content serving as a secondary engagement layer to broaden appeal and maximize tourism-intention conversion.

6. Conclusions

6.1. Summary of Findings

This study examines Chinese tourists' engagement with Lamphun's intangible cultural heritage through an integrated digital analytics framework, with the aim of identifying sentiment patterns, tourism-intention classification, and thematic trends. The findings revealed a predominance of positive sentiment toward Lamphun's cultural offerings, indicating a favorable perception of cultural tourism resources, with the highest proportion of negative sentiment in 'Knowledge and practices concerning nature and the universe', indicating that culinary-related expectations among Chinese visitors present the most significant area for destination improvement.

Positive sentiment significantly associated with tourism-intention signal posts with sentiment score emerged as the strongest predictor of tourism intention in the Random Forest classification model, and platform-normalized engagement metrics served as supplementary predictors. Platform-stratified analysis demonstrated that the sentiment-intention relationship was significant, specifically on Rednote, confirming the platform's distinctive role as a travel inspiration and decision-influencing channel for Chinese tourists. Temporal stratification further revealed that the sentiment-intention association was most pronounced during the pandemic restriction period, consistent with compensatory travel motivation behaviors.

Three thematic clusters were identified, encompassing culinary and lifestyle experiences, hybrid cafe-culture and religious-site content, and festival and ritual symbolism. 'Cafe exploration and religious-related site' cluster achieved the highest positive sentiment distribution and tourism-intention rate. Cluster 1 reflected everyday culinary and lifestyle exploration and generated the broadest digital reach but the lowest tourism-intention rate. Cluster 2, combining cafe culture, local cuisine, and religious-related sites, produced the highest positive sentiment distribution and the highest tourism-intention rate. Cluster 3, centered on festival and ritual symbolism, demonstrated strong affective engagement and a notably higher tourism-intention rate than the culinary lifestyle cluster. Collectively, these findings indicate that festival-centered and hybrid heritage—lifestyle content represent the most strategically valuable content categories for driving Chinese tourist engagement and travel consideration toward Lamphun.

6.2. Theoretical Implications

This study makes several contributions to the theoretical literature on digital heritage tourism and UGC analytics. First, the findings advance the understanding of destination

image formation in the context of lesser-known ICH destinations, demonstrating that the digital image of Lamphun among Chinese tourists is multidimensional and co-constructed across experiential themes encompassing cultural festivals, religious heritage, culinary traditions, and contemporary lifestyle activities. This extends prior work on destination image formation through social media by applying this framework to a secondary, under-represented heritage context with low baseline digital visibility.

Second, the study contributes to the growing body of research on Chinese digital tourism behavior by providing platform-differentiated empirical evidence of the sentiment-intention relationship across WeChat, Douyin, and Rednote. The platform-specific findings reinforce the theoretical position that distinct digital ecosystems mediate tourism decision-making in structurally different ways and extend this argument by demonstrating that the strength of the sentiment-intention association varies significantly by platform type.

Third, the integrated multi-method framework combining sentiment analysis, rule-based tourism intention classification, Random Forest modeling, TF-IDF analysis, and K-means clustering offers a methodological contribution to tourism analytics research. The framework demonstrates that analytical approaches combining affective, behavioral, and thematic dimensions of UGC provide richer and more actionable insights than single-method approaches and that this integrated approach is feasible and informative even in data-limited contexts characteristic of lesser-known destinations.

6.3. Managerial Implications

The findings carry concrete implications for DMOs and cultural heritage managers responsible for promoting Lamphun as a tourism destination. The integration of contemporary cafe culture and local cuisine narratives with religious-related site content should be considered as co-branding strategies that present Lamphun as a culturally immersive yet lifestyle-compatible destination, particularly targeting younger Chinese travelers whose preferences reflect a diversification of travel motivations beyond traditional sightseeing. As culinary culture represents the most discussed category, DMOs should develop coordinated, visually rich campaign materials oriented around these events and distribute proactively. In this context of Chinese visitors and Lamphun, Rednote should be prioritized as the primary channel for destination content seeding and influencer partnerships, while WeChat and Douyin serve complementary functions in awareness-building and community engagement.

The strong co-occurrence of Chiang Mai indicates that Chinese tourists currently perceive Lamphun primarily as a secondary stop within a Chiang Mai-centered itinerary. DMOs in Lamphun should therefore pursue collaborative regional marketing arrangements with Chiang Mai counterparts, developing joint content and itinerary packages that leverage the established pull of Chiang Mai to drive purposeful secondary visitation to Lamphun. This positioning strategy is more practically achievable than repositioning Lamphun as an independent primary destination in the short term.

6.4. Limitations and Future Research

Several limitations of this study should be acknowledged. First, the dataset was confined to three Chinese social media platforms between 2019 and 2023, with post volume concentrated disproportionately in the pandemic restriction years and on WeChat. Consequently, the findings should be interpreted with caution, given this uneven temporal and platform distribution. Second, although the rule-based tourism intention classifier was validated through inter-coder reliability assessment and comparison against a gold standard subset, it achieved a recall of 0.600 for the intention class, meaning a meaningful share of intention-bearing posts may not have been captured. Third, this study focused exclusively

on Chinese tourists. While this specificity was theoretically motivated by the distinctive characteristics of the Chinese digital ecosystem, it constrains the generalizability of the findings to other international visitor segments. Finally, machine translation tools were used to convert Chinese text into English for the TF IDF and thematic clustering analyses. Although this supports accessibility for an international readership, translation-induced semantic shifts cannot be fully excluded.

Future studies should incorporate comparative analyses across international tourist segments to provide a more comprehensive understanding of destination perception and engagement at Lamphun and other underrepresented destinations. This framework could also be extended to other underrepresented ICH destinations, enabling cross-destination comparison and supporting the development of a generalized UGC analytics model for secondary cultural heritage tourism marketing. Finally, integrating generative AI tools for automated content strategy recommendations based on UGC analytics outputs represents a particularly promising avenue for bridging empirical research and actionable destination management organization (DMO) practice.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/tourhosp7070186/s1>, Table S1: Methodological mapping of literature-retrieved intention indicator keywords and localized script classification tokens. References (Gregoriades et al., 2023; F. Ye et al., 2025) are cited in the supplementary materials.

Author Contributions: Conceptualization, P.W. and N.M.; Methodology, P.W., P.O. and C.K.; Software, P.W.; Validation, P.W., P.O. and N.M.; Formal Analysis, P.W.; Investigation, P.W.; Data Curation, P.W.; Writing—Original Draft Preparation, P.W.; Writing—Review and Editing, P.W., P.O., C.K. and N.M.; Visualization, P.W.; Supervision, P.O. and C.K.; Project Administration, P.W. and P.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors wish to express their sincere gratitude to Nopasit Chakpitak and Aniwat Phaphuangwittayakul for their valuable guidance and suggestions on machine learning methodologies employed in this study. The authors also acknowledge the International College of Digital Innovation (ICDI), Chiang Mai University, for providing access to computational resources and facilities that supported the machine learning components of this research. During the preparation of this manuscript, the authors used AI-assisted machine translation tools to support the preprocessing of Chinese-language source data, as described in Section 2.3. AI language tools were additionally used to support grammar correction and academic language editing. All intellectual content, data, analysis, interpretations, and conclusions are the authors' own. The authors have reviewed and edited the outputs and take full responsibility for the content of this publication.

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

References

- Abdelhamed, M. (2023). Digital transformation in tourism and its role in preserving heritage. *Journal of Tourism, Hotels and Heritage*, 7(1), 44–57. [CrossRef]
- Abdo, A. (2019). Digital heritage applications and its impact on cultural tourism. *Journal of Association of Arab Universities for Tourism and Hospitality*, 17(1), 37–50. [CrossRef]
- Abel, J., & Lantow, B. (2019). A methodological framework for dictionary and rule-based text classification. In *Proceedings of the 11th international joint conference on knowledge discovery, knowledge engineering and knowledge management (IC3K 2019)*, Vienna, Austria, September 17–19 (pp. 330–337). SciTePress. [CrossRef]

- Afshardoost, M., & Eshaghi, M. S. (2020). Destination image and tourist behavioural intentions: A meta-analysis. *Tourism Management*, 81, 104154. [CrossRef]
- Al-Ababneh, M. A., & Masadeh, M. (2019). Creative cultural tourism as a new model for cultural tourism. *Journal of Tourism Management Research*, 6(2), 109–118. [CrossRef]
- Alivizatou, M. (2012). *Intangible heritage and the museum: New perspectives on cultural preservation*. Left Coast Press.
- Aljarah, A., Sawaftah, D., Ibrahim, B., & Lahuerta-Otero, E. (2022). The differential impact of user- and firm-generated content on online brand advocacy: Customer engagement and brand familiarity matter. *European Journal of Innovation Management*, 27(4), 1160–1181. [CrossRef]
- Angelini, L., Borlizzi, D., Carlucci, A., Ciardella, G., Destefanis, A., Governale, G., & Morfini, I. (2020). Cultural tourism development and the impact on local communities: A case study from the south of Italy. *CERN IdeaSquare Journal of Experimental Innovation*, 4(2), 19–24. [CrossRef]
- Arefieva, V., Egger, R., & Yu, J. (2021). A machine learning approach to cluster destination image on Instagram. *Tourism Management*, 85, 104318. [CrossRef]
- Asyraf, M. A., Hanafiah, M. H., Zain, N. a. M., & Hariani, D. (2023). Unboxing the paradox of social media user-generated content (UGC) information qualities and tourist behaviour: Moderating effect of perceived travel risk. *Journal of Hospitality and Tourism Insights*, 7(4), 1809–1830. [CrossRef]
- Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., Pouliquen, B., & Belyaeva, J. (2013). Sentiment analysis in the news. *arXiv*. [CrossRef]
- Basnyat, S., & Jiahui, Y. (2020). The role of social media in shaping tourists' travel decisions. In *Handbook of research on social media applications for the tourism and hospitality sector* (pp. 52–65). IGI Global Scientific Publishing. [CrossRef]
- Bathla, G., Raina, A., Tripathi, R., Rana, V. S., & Singh, A. (2024). Assessing the impact of digital innovations on tourist decision making and behavior. In *Tracking tourism patterns and improving travel experiences with innovative technologies* (pp. 317–334). IGI Global Scientific Publishing. [CrossRef]
- Bayerl, P. S., & Paul, K. I. (2011). What determines inter-coder agreement in manual annotations? A meta-analytic investigation. *Computational Linguistics*, 37(4), 699–725. [CrossRef]
- Bigne, E., Ruiz, C., & Curras-Perez, R. (2019). Destination appeal through digitalized comments. *Journal of Business Research*, 101, 447–453. [CrossRef]
- Borrajó-Millán, F., Alonso-Almeida, M., Escat-Cortes, M., & Yi, L. (2021). Sentiment analysis to measure quality and build sustainability in tourism destinations. *Sustainability*, 13(11), 6015. [CrossRef]
- Boukhari, M. E., Amajid, G., & Fadili, M. H. (2025). Tourism in the age of artificial intelligence: A conceptual perspective on marketing and beyond. *IJDAM—International Journal of Digitalization and Applied Management*, 2(2), 117–131. [CrossRef]
- Chemin, M., Da Silva, C. P., & De Paul Vikou, S. V. (2025). User-generated content (UGC) in tourist attractions and destinations: Systematic literature review and perspectives for management. *PASOS Revista de Turismo y Patrimonio Cultural*, 23(2), 539–562. [CrossRef]
- Chen, J. (2025). Research on the digital transformation strategy and organizational structure optimization of intangible cultural heritage. *Modern Economics & Management Forum*, 5(6), 1010. [CrossRef]
- Chendraningrum, D., & Suryana, M. (2025). Narrating the destination through food: An empirical study of cultural cuisine, tourist engagement, and attractiveness. *Journal of Tourism Sustainability*, 5(1), 1–14. [CrossRef]
- Cheng, J., & Jiang, S. (2024). Exploring Thailand as a gastronomic tourism destination: A structural equation model approach. *Journal of Infrastructure Policy and Development*, 8(11), 8214. [CrossRef]
- Cheng, L., Chen, X., De Vos, J., Lai, X., & Witlox, F. (2018). Applying a random forest method approach to model travel mode choice behavior. *Travel Behaviour and Society*, 14, 1–10. [CrossRef]
- Cheung, M. L., Leung, W. K., Cheah, J., & Ting, H. (2021). Exploring the effectiveness of emotional and rational user-generated contents in digital tourism platforms. *Journal of Vacation Marketing*, 28(2), 152–170. [CrossRef]
- Chhabra, D., Healy, R., & Sills, E. (2003). Staged authenticity and heritage tourism. *Annals of Tourism Research*, 30(3), 702–719. [CrossRef]
- China Internet Network Information Center. (2023). *The 51st statistical report on China's internet development*. China Internet Network Information Center (CNNIC). Available online: <https://www.cnnic.com.cn/IDR/ReportDownloads/202307/P020230707514088128694.pdf> (accessed on 1 June 2023).
- Choi, S., Lehto, X. Y., & Morrison, A. M. (2007). Destination image representation on the web: Content analysis of Macau travel related websites. *Tourism Management*, 28(1), 118–129. [CrossRef]
- Dai, P. (2023). The evolution of Chinese internet culture: A study on the social media platforms' role and their impact on online trends. *Communications in Humanities Research*, 12(1), 254–262. [CrossRef]
- Deng, M. (2025). Machine learning advances in technology applications: Cultural heritage tourism trends in experience design. *International Journal of Advanced Computer Science and Applications*, 16(4), 186. [CrossRef]

- Department of Cultural Promotion in Thailand. (n.d.). *Intangible cultural heritage of Thailand*. Intangible Cultural Heritage of Thailand. Available online: <https://ich-thailand.org/> (accessed on 20 May 2023).
- Di Gangi, P. M., & Wasko, M. M. (2016). Social media engagement theory. *Journal of Organizational and End User Computing*, 28(2), 53–73. [CrossRef]
- Dligach, D., Nielsen, R. D., & Palmer, M. (2010). To annotate more accurately or to annotate more. In *Proceedings of the fourth linguistic annotation workshop, ACL 2010, Uppsala, Sweden, July 15–16* (pp. 64–72). Association for Computational Linguistics. Available online: <https://digital.library.unt.edu/ark:/67531/metadc1042573/> (accessed on 21 June 2026).
- Dobrea, R., Marin, A., Dima, C., & Moncea, M. (2023). The relationship between the tourism industry and sustainable development goals word cloud analysis. *Amfiteatru Economic*, 25(17), 1131. [CrossRef]
- Espeso-Molinero, P. (2022). Trends in cultural tourism. *Smart Tourism*, 3(2), 13. [CrossRef]
- Fakfare, P., Talawanich, S., & Wattanacharoensil, W. (2020). A scale development and validation on domestic tourists' motivation: The case of second-tier tourism destinations. *Asia Pacific Journal of Tourism Research*, 25(5), 489–504. [CrossRef]
- Filatova, O. (2016). More than a word cloud. *TESOL Journal*, 7(2), 438–448. [CrossRef]
- Fotis, J., Buhalis, D., & Rossides, N. (2012). Social media use and impact during the holiday travel planning process. In *Information and communication technologies in tourism 2012* (pp. 13–24). Springer. [CrossRef]
- Galani, S., & Vosinakis, S. (2024). An augmented reality approach for communicating intangible and architectural heritage through digital characters and scale models. *Personal and Ubiquitous Computing*, 28(3–4), 471–490. [CrossRef]
- Gao, J., Zhang, C., Zhou, X., & Cao, R. (2021). Chinese tourists' perceptions and consumption of cultural heritage: A generational perspective. *Asia Pacific Journal of Tourism Research*, 26(7), 719–731. [CrossRef]
- Gau, L., Chu, H., Pham, D. T., & Huang, C. (2024). Innovative teaching of AI-based text mining and ChatGPT applications for trend recognition in tourism and hospitality. *Tourism and Hospitality*, 5(4), 1274–1291. [CrossRef]
- Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Data mining and analytics in the process industry: The role of machine learning. *IEEE Access*, 5, 20590–20616. [CrossRef]
- Gnoth, J., & Zins, A. H. (2011). Developing a tourism cultural contact scale. *Journal of Business Research*, 66(6), 738–744. [CrossRef]
- Gour, A., Aggarwal, S., & Erdem, M. (2021). Reading between the lines: Analyzing online reviews by using a multi-method web-analytics approach. *International Journal of Contemporary Hospitality Management*, 33(2), 490–512. [CrossRef]
- Govers, R., Go, F. M., & Kumar, K. (2007). Promoting tourism destination image. *Journal of Travel Research*, 46(1), 15–23. [CrossRef]
- Gregoriades, A., Pampaka, M., Herodotou, H., & Christodoulou, E. (2023). Explaining tourist revisit intention using natural language processing and classification techniques. *Journal of Big Data*, 10(1), 60. [CrossRef]
- Hagenauer, J., & Helbich, M. (2017). A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*, 78, 273–282. [CrossRef]
- Hamroun, M., & Gouider, M. S. (2020). A survey on intention analysis: Successful approaches and open challenges. *Journal of Intelligent Information Systems*, 55(3), 423–443. [CrossRef]
- Harish, T., Neeha, S., Deepika, T., Haritha, L., & Kameswari, T. (2023). Machine learning approach to cluster destination image on social networks. *International Journal of Innovative Research in Engineering & Management*, 10(2), 118–120. [CrossRef]
- He, H. (2023). Analysis of cultural and tourism promotion in Chinese small and medium-sized cities on new media platforms. *SHS Web of Conferences*, 179, 03011. [CrossRef]
- Hou, X., & Paidi, Z. B. (2025). AI-based tourist behavior analysis and cultural communication optimization strategies for Shanxi great wall heritage site. *Future Technology*, 4(4), 43–58. [CrossRef]
- Hou, Y., Kenderdine, S., Picca, D., Egloff, M., & Adamou, A. (2022). Digitizing intangible cultural heritage embodied: State of the art. *Journal on Computing and Cultural Heritage*, 15(3), 1–20. [CrossRef]
- Hu, T., & Chen, H. (2023). Little-known leisure places: Chinese tourists' preferences for visiting niche tourism destinations. *Asia Pacific Journal of Tourism Research*, 28(11), 1261–1278. [CrossRef]
- Hua, Y., Ding, L., Dong, H., & Lin, Z. (2024). Influence of user-generated content (UGC) in social media on the intangible cultural heritage preservation of Gen Z tourists in the digital economy era. *International Journal of Tourism Research*, 26(5), e2743. [CrossRef]
- Huang, A., De La Mora Velasco, E., Haney, A., & Alvarez, S. (2022). The future of destination marketing organizations in the insight era. *Tourism and Hospitality*, 3(3), 803–808. [CrossRef]
- Huang, Y. (2024). Multimodal communication and city branding on the Xiaohongshu platform: An analysis of Quanzhou's digital heritage tourism. *Journal of Humanities Arts and Social Science*, 8(10), 2313–2318. [CrossRef]
- Islam, M. T. (2021). Applications of social media in the tourism industry: A review. *SEISENSE Journal of Management*, 4(1), 59–68. [CrossRef]
- Iswanto, D., & Irsyad, Z. (2024). Culture as a tourism magnet a systematic review of the influence of local traditions on visitor engagement in tourism destinations. *Journal of Economic Growth and Development Review*, 1, 35–44. [CrossRef]
- Jain, A. K. (2009). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. [CrossRef]

- Kaewyu, P., Pakdeepinit, P., & Madtyampurush, W. (2020). Behaviour of independent Chinese tourists travelling in Lanna civilization area, Thailand. *WMS Journal of Management*, 9(3), 102–112.
- Kaukuntla, P. R. (2025). Analyzing the impact of social proof and user-generated content on engagement using data-driven methods. *International Scientific Journal of Engineering and Management*, 4(2), 1–7. [\[CrossRef\]](#)
- Kiatkawsin, K., & Han, H. (2017). An alternative interpretation of attitude and extension of the value–attitude–behavior hierarchy: The destination attributes of Chiang Mai, Thailand. *Asia Pacific Journal of Tourism Research*, 22(5), 481–500. [\[CrossRef\]](#)
- Kim, E. E. K., Seo, K., & Choi, Y. (2021). Compensatory travel post COVID-19: Cognitive and emotional effects of risk perception. *Journal of Travel Research*, 61(8), 1895–1909. [\[CrossRef\]](#)
- Konak, Y. (2024). The impact of user-generated content on social media marketing effectiveness. *Marketing and Digital Technologies*, 7(3), 83–98. [\[CrossRef\]](#)
- Kwale, F. M. (2013). A critical review of K means text clustering algorithms. *International Journal of Advanced Research in Computer Science*, 4(9), 27–34.
- Lam, T., & Hsu, C. H. (2005). Predicting behavioral intention of choosing a travel destination. *Tourism Management*, 27(4), 589–599. [\[CrossRef\]](#)
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159. [\[CrossRef\]](#)
- Leelawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtabtim, K., Laosunthara, A., Yudha, A. K., & Tang, J. (2022). Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning. *Heliyon*, 8(10), e10894. [\[CrossRef\]](#) [\[PubMed\]](#)
- Leung, D., Law, R., Van Hoof, H., & Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of Travel & Tourism Marketing*, 30(1–2), 3–22. [\[CrossRef\]](#)
- Li, H., Pan, J., Hui, Y., Liu, J., & Nijkamp, P. (2023). The geography of social media platform attention for tourist attractions-spatial digital data analytics of scenic hot spots in China. *Eastern Journal of European Studies*, 14, 135–158. [\[CrossRef\]](#)
- Li, J., Peng, X., Liu, X., Tang, H., & Li, W. (2024). A study on shaping tourists' conservational intentions towards cultural heritage in the digital era: Exploring the effects of authenticity, cultural experience, and place attachment. *Journal of Asian Architecture and Building Engineering*, 24(3), 1965–1984. [\[CrossRef\]](#)
- Li, L., Zheng, S., & Wang, Z. (2015). An exploratory study on social media in China. In *Information and communication technologies in tourism 2015* (pp. 255–267). Springer. [\[CrossRef\]](#)
- Li, N., Liu, X., & Meng, F. (2025). Machine learning techniques in tourism and hospitality research: A critical assessment. *Journal of Travel Research*. [\[CrossRef\]](#)
- Li, Q., Li, S., Zhang, S., Hu, J., Hu, J., Hu, J., & Hu, J. (2019). A review of text corpus-based tourism big data mining. *Applied Sciences*, 9(16), 3300. [\[CrossRef\]](#)
- Li, Y., Lin, Z., & Xiao, S. (2022). Using social media big data for tourist demand forecasting: A new machine learning analytical approach. *Journal of Digital Economy*, 1(1), 32–43. [\[CrossRef\]](#)
- Li, Z., Jia, L., & Su, B. (2019). Improved K-means algorithm for finding public opinion of mount Emei tourism. In *Proceedings of the 2019 15th international conference on computational intelligence and security (CIS), Macao, China, 13–16 December* (pp. 192–196). IEEE. [\[CrossRef\]](#)
- Lin, C. Y., & Li, C. (2024). Digitization of intangible cultural heritage as a key factor in the transformation of cultural tourism integration: An analytical perspective of the digitalization framework. *Trends in Social Sciences and Humanities Research*, 2(1), 22–40. [\[CrossRef\]](#)
- Lin, M. S., Liang, Y., Xue, J. X., Pan, B., & Schroeder, A. (2021). Destination image through social media analytics and survey method. *International Journal of Contemporary Hospitality Management*, 33(6), 2219–2238. [\[CrossRef\]](#)
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458–468. [\[CrossRef\]](#)
- Liu, B. (2012). Sentiment analysis and opinion mining. In *Synthesis lectures on human language technologies* (Vol. 5, pp. 1–167). Springer. [\[CrossRef\]](#)
- Liu, J. (2024). Research on trending algorithms of digital platforms: A case study of Little Red Book and Instagram. *Media and Communication Research*, 5(4), 58–68. [\[CrossRef\]](#)
- Liu, J., Hu, S., Mehraliyev, F., & Liu, H. (2023). Text classification in tourism and hospitality—A deep learning perspective. *International Journal of Contemporary Hospitality Management*, 35(12), 4177–4190. [\[CrossRef\]](#)
- Liu, W., An, G., Xue, Y., & Hong, C. (2025). Multimodal analysis of the destination image of traditional villages based on tourist comments and photos: The case of Hongcun, China. *Journal of Hospitality and Tourism Technology*, 17(2), 561–581. [\[CrossRef\]](#)
- Liu, X., Mehraliyev, F., Liu, C., & Schuckert, M. (2019). The roles of social media in tourists' choices of travel components. *Tourist Studies*, 20(1), 27–48. [\[CrossRef\]](#)
- Liu, Y., & Beldona, S. (2021). Extracting revisit intentions from social media big data: A rule-based classification model. *International Journal of Contemporary Hospitality Management*, 33(6), 2176–2193. [\[CrossRef\]](#)

- Lo, M., Chin, C., & Law, F. (2019). Tourists' perspectives on hard and soft services toward rural tourism destination competitiveness: Community support as a moderator. *Tourism and Hospitality Research*, 19(2), 139–157. [CrossRef]
- Lojo, A. (2020). Young Chinese in Europe. *Tourism*, 68(1), 7–20. [CrossRef]
- López-Guzmán, T., José, M., & Elide, D. (2016). Culinary travel as new approach for cultural tourism. *Turizam*, 20(1), 1–11. [CrossRef]
- Lushchik, M. V. (2023). Features of the outbound tourist market of China in the current realities. *Journal of Geology Geography and Geoecology*, 32(2), 310–325. [CrossRef] [PubMed]
- Ma, J. X., Buhalis, D., & Song, H. (2003). ICTs and internet adoption in China's tourism industry. *International Journal of Information Management*, 23(6), 451–467. [CrossRef]
- Ma, S. N. Z., & Mohame, B. B. (2023). Heritage tourism and cultural identity in China: Challenges and opportunities for development. *International Journal of Professional Business Review*, 8(7), e02393. [CrossRef]
- Mariani, M. M., Di Felice, M., & Mura, M. (2016). Facebook as a destination marketing tool: Evidence from Italian regional destination management organizations. *Tourism Management*, 54, 321–343. [CrossRef]
- Marine-Roig, E., & Clavé, S. A. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, 4(3), 162–172. [CrossRef]
- Mathur, S., Tewari, A., & Singh, A. (2021). Modeling the factors affecting online purchase intention: The mediating effect of consumer's attitude towards user-generated content. *Journal of Marketing Communications*, 28(7), 725–744. [CrossRef]
- Ministry of Tourism and Sports. (2023). *Tourism statistics 2023*. Ministry of Tourism and Sports. Available online: <https://www.mots.go.th/news/category/704> (accessed on 14 December 2023).
- Mirzaalian, F., & Halpenny, E. (2019). Social media analytics in hospitality and tourism. *Journal of Hospitality and Tourism Technology*, 10(4), 764–790. [CrossRef]
- Mkono, M. (2012). Authenticity does matter. *Annals of Tourism Research*, 39(1), 480–483. [CrossRef]
- Mo, L., Aris, A. B., & Yan, X. (2024). Digital display and storytelling: Creating an immersive experience of Lingnan costume cultural heritage tourism project. *IOP Conference Series Earth and Environmental Science*, 1366(1), 012047. [CrossRef]
- Neuhofer, B., Buhalis, D., & Ladkin, A. (2015). Smart technologies for personalized experiences: A case study in the hospitality domain. *Electronic Markets*, 25(3), 243–254. [CrossRef]
- Ngai, E. W., Tao, S. S., & Moon, K. K. (2014). Social media research: Theories, constructs, and conceptual frameworks. *International Journal of Information Management*, 35(1), 33–44. [CrossRef]
- Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., García, Á. L., Heredia, I., Malík, P., & Hluchý, L. (2019). Machine learning and deep learning frameworks and libraries for large-scale data mining: A survey. *Artificial Intelligence Review*, 52(1), 77–124. [CrossRef]
- Nicolau, J. L. (2025). When ChatGPT designs your trip: How GenAI adds a cognitive layer to smart tourism. *Journal of Smart Tourism*, 5(4), 137–151. [CrossRef]
- Nikolskaya, E. Y., Zakharova, E. V., Galkin, D. V., Kovaleva, N. I., & Panova, N. A. (2021). The impact of digital technologies on the transformation of the tourism and hospitality industry. *Revista Gestão Inovação e Tecnologias*, 11(4), 623–632. [CrossRef]
- Pang, H., Wang, J., & Hu, X. (2021). Understanding the potential influence of WeChat engagement on bonding capital, bridging capital, and electronic word-of-mouth intention. *Sustainability*, 13(15), 8489. [CrossRef]
- Qiu, Q., & Zhang, M. (2021). Using content analysis to probe the cognitive image of intangible cultural heritage tourism: An exploration of Chinese social media. *ISPRS International Journal of Geo-Information*, 10(4), 240. [CrossRef]
- Qiu, Q., Zuo, Y., & Zhang, M. (2022). Intangible cultural heritage in tourism: Research review and investigation of future agenda. *Land*, 11(1), 139. [CrossRef]
- Quer, D., & Peng, J. (2022). Development of China's outbound tourism. In *Chinese outbound tourist behaviour* (pp. 61–73). Routledge. [CrossRef]
- Rahmani, A. M., Azhir, E., Ali, S., Mohammadi, M., Ahmed, O. H., Ghafour, M. Y., Ahmed, S. H., & Hosseinzadeh, M. (2021). Artificial intelligence approaches and mechanisms for big data analytics: A systematic study. *PeerJ Computer Science*, 7, e488. [CrossRef] [PubMed]
- Richa, R., & Naithani, B. P. (2024). A review of cultural tourism. *Journal of Advances and Scholarly Researches in Allied Education*, 21(2), 45–50. [CrossRef]
- Richards, G. (2018). Cultural tourism: A review of recent research and trends. *Journal of Hospitality and Tourism Management*, 36, 12–21. [CrossRef]
- Sanjaya, V., & Aruan, T. D. H. (2021). The link between travel motivation and satisfaction in heritage destination: The role of experiential factors and heritage destination image. *IOP Conference Series Earth and Environmental Science*, 716(1), 012067. [CrossRef]
- Satchapappichit, S. (2020). Factors influencing Chinese visitors' dining experiences with Thai cuisine in Bangkok, Thailand. *ABAC Journal*, 40(4), 58–77.
- Sawant, P. G., & Desai, B. L. (2015). Big data mining: Challenges and opportunities to forecast future scenario. *International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)*, 3(6), 5228–5232.

- Seyfi, S., Hall, C. M., & Rasoolimanesh, S. M. (2019). Exploring memorable cultural tourism experiences. *Journal of Heritage Tourism*, 15(3), 341–357. [CrossRef]
- Shao, Z., & Li, Y. (2025). Research and Practice of artificial intelligence in improving the efficiency and effect of digital communication of intangible cultural heritage—A case study of Yungang culture. *Journal of Posthumanism*, 5(3), 953–979. [CrossRef]
- Sharma, I., & Aggarwal, A. (2024). Exploring the relationship between tourist experiences at creative tourist destinations and their intentions to indulge in co-creation activities. In *Advances in hospitality, tourism and the services industry (AHTSI) book series* (pp. 71–87). IGI Global. [CrossRef]
- Shu, X., & Ye, Y. (2022). Knowledge discovery: Methods from data mining and machine learning. *Social Science Research*, 110, 102817. [CrossRef] [PubMed]
- Solazzo, G., Maruccia, Y., Lorenzo, G., Ndou, V., Del Vecchio, P., & Elia, G. (2021). Extracting insights from big social data for smarter tourism destination management. *Measuring Business Excellence*, 26(1), 122–140. [CrossRef]
- Stepchenkova, S., & Morrison, A. M. (2008). Russia's destination image among American pleasure travelers: Revisiting Echtner and Ritchie. *Tourism Management*, 29(3), 548–560. [CrossRef]
- Sujatmiko, S., Ar, D. P., Hamdat, A., & Salam, K. N. (2025). User-generated content (UGC) and its impact on tourism marketing: A systematic literature review. *Golden Ratio of Mapping Idea and Literature Format*, 5(2), 97–105. [CrossRef]
- Tan, Y. (2026). Non-place making: Chinese tourists' production of tourist spaces through social media. *Tourism Geographies*, 28(3), 427–451. [CrossRef]
- Tang, L., Omar, S. Z., Bolong, J., & Zawawi, J. W. M. (2021). Social media use among young people in China: A systematic literature review. *SAGE Open*, 11(2). [CrossRef]
- Theocharis, D., Tsekouropoulos, G., Chatzigeorgiou, C., & Kokkinis, G. (2025). Empirical categorization of factors affecting online consumer behavior of Gen Z regarding newly launched technological products and moderating impact of perceived risk. *Behavioral Sciences*, 15(3), 371. [CrossRef] [PubMed]
- Tian, L., Pu, W., Su, C., Chen, M., & Lin, Y. (2021). Asymmetric effects of China's tourism on the economy at the city level: A moderating role of spatial disparities in top level tourist attractions. *Current Issues in Tourism*, 25(16), 2648–2664. [CrossRef]
- Tourism Office, Lamphun City Hall. (n.d.). *Lamphun destinations experience the charm of a timeless cultural city*. Tourism in Lamphun. Available online: <https://travel.lamphunpao.go.th/en> (accessed on 3 June 2023).
- Tri, T. D., Hung, H. T., La Phuong Hien, H., Vu, L. N. A., & Hao, D. D. Q. (2023). Tourists' engagement behaviors on social media, destination advocacy and revisit intention. *Hue University Journal of Science Economics and Development*, 132(5B), 99–112. [CrossRef]
- Tsekouropoulos, G. (2019). Viral advertising: Message quality, trust and consumers intention to share the content in social media. *International Journal of Technology Marketing*, 13(2), 111. [CrossRef]
- Tu, W., Zhou, L., Haobin, B. Y., & Yan, Q. (2022). Conceptualizing and assessing the competitiveness of slow tourism destinations: Evidence from the first accredited Cittaslow in China. *SAGE Open*, 12(1), 1–12. [CrossRef]
- UNESCO. (2003). *ICH—Intangible cultural heritage*. UNESCO—Intangible Cultural Heritage. Available online: <https://ich.unesco.org/en/home> (accessed on 6 May 2023).
- Wang, D., Xiang, Z., & Fesenmaier, D. R. (2014). Smartphone use in everyday life and travel. *Journal of Travel Research*, 55(1), 52–63. [CrossRef]
- Wang, J., & Xia, L. (2021). Revenge travel: Nostalgia and desire for leisure travel post COVID-19. *Journal of Travel & Tourism Marketing*, 38(9), 935–955. [CrossRef]
- Wang, T., & Park, J. (2023). Destination information search in social media and travel intention of Generation Z university students. *Journal of China Tourism Research*, 19(3), 570–588. [CrossRef]
- Wang, W., Chen, J. S., & Huang, K. (2015). Religious tourist motivation in Buddhist mountain: The case from China. *Asia Pacific Journal of Tourism Research*, 21(1), 57–72. [CrossRef]
- Wang, Z., Huang, W., & Liu-Lastres, B. (2022). Impact of user-generated travel posts on travel decisions: A comparative study on Weibo and Xiaohongshu. *Annals of Tourism Research Empirical Insights*, 3(2), 100064. [CrossRef]
- Wanju, T., Zijuan, W., & Yijie, T. (2025). ICH and e-commerce: Bridging tradition and modernity in cultural and creative products. *Journal of Economics Management and Trade*, 31(1), 1–8. [CrossRef]
- Wasela, K. (2023). The role of intangible cultural heritage in the development of cultural tourism. *International Journal of Eco-Cultural Tourism, Hospitality Planning and Development*, 6(2), 15–28. [CrossRef]
- World Tourism Organization. (2019). UNWTO tourism definitions. In *World Tourism Organization (UNWTO) eBooks*. UNWTO. [CrossRef]
- Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, 31(2), 179–188. [CrossRef]
- Xu, H., Cheung, L. T., Lovett, J., Duan, X., Pei, Q., & Liang, D. (2021). Understanding the influence of user-generated content on tourist loyalty behavior in a cultural World Heritage Site. *Tourism Recreation Research*, 48(2), 173–187. [CrossRef]
- Xu, Z., Liu, F., Xia, G., Wang, S., Duan, Y., Yu, L., Zhao, S., & Li, M. (2025). Immersive HCI for intangible cultural heritage in tourism contexts: A narrative review of design and evaluation. *Sustainability*, 18(1), 153. [CrossRef]

- Yan, G., He, W., Shen, J., & Tang, C. (2014). A bilingual approach for conducting Chinese and English social media sentiment analysis. *Computer Networks*, 75, 491–503. [\[CrossRef\]](#)
- Yan, Q., Jiang, T., Zhou, S., & Zhang, X. (2022). Exploring tourist interaction from user-generated content: Topic analysis and content analysis. *Journal of Vacation Marketing*, 30(2), 327–344. [\[CrossRef\]](#)
- Yang, Y., & Shen, H. (2024). Digital marketing practices in tourism: Advances in the cultural heritage of China. In *Handbook of experience science* (pp. 222–236). Edward Elgar Publishing. [\[CrossRef\]](#)
- Ye, F., Yin, M., Sun, S., & Wang, X. (2025). From Watching to Wishing: A multimodal computational analysis of how PUGC Video-Danmaku ecology shapes travel intention. *Information Technology & Tourism*, 28(1), 39. [\[CrossRef\]](#)
- Ye, S., Shi, L., Wu, J., & Hyuk, G. (2025). Beyond the tangible: Factors influencing memorable experiences in intangible culture heritage tourism in Wuxi, China. *SAGE Open*, 15(1), 1–18. [\[CrossRef\]](#)
- Yuan, Y., Chan, C., Eichelberger, S., Ma, H., & Pikkemaat, B. (2026). The effect of social media on travel planning process by Chinese tourists: The way forward to tourism futures. *Journal of Tourism Futures*, 12, 121–140. [\[CrossRef\]](#)
- Zayachka, P. (2023). Intangible cultural heritage as a resource for development of sustainable cultural tourism. *Cultural and Historical Heritage Preservation Presentation Digitalization*, 9(1), 141–153. [\[CrossRef\]](#)
- Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives*, 10, 27–36. [\[CrossRef\]](#)
- Zhang, B., & Niyomsilp, E. (2020). The relationship between tourism destination image, perceived value and post-visiting behavioral intention of Chinese tourist to Thailand. *International Business Research*, 13(11), 96. [\[CrossRef\]](#)
- Zhang, Y. (2025). Exploring the value reshaping and practical approaches to the digital inheritance of cultural heritage from the perspective of culture-tourism integration. *Educational Innovation Research*, 3(4), 205–211. [\[CrossRef\]](#)
- Zhang, Y., & Lee, T. J. (2021). Alienation and authenticity in intangible cultural heritage tourism production. *International Journal of Tourism Research*, 24(1), 18–32. [\[CrossRef\]](#)
- Zhang, Y., Lingyi, M., Peixue, L., Lu, Y., & Zhang, J. (2021). COVID-19's impact on tourism: Will compensatory travel intention appear? *Asia Pacific Journal of Tourism Research*, 26(7), 732–747. [\[CrossRef\]](#)
- Zou, M., Lei, M., & Fan, Q. (2025). XiaoHongshu platform usage: Exploring the impact of content values, user satisfaction, and continuance intention among Chinese tourists. *Advances in Information Management and Economic Development Research*, 2(1), 53. [\[CrossRef\]](#)

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.