Human Factors in Communication of Design, Vol. 177, 2025, 89–98
https://doi.org/10.54941/ahfe1006334

From Data to Insights: Perceiving Museum Experiences Through Natural Language Processing of Social Media Corpus

Xueqi Liu[1] and Zhiping Qu[2]

[1]Hebei Jiaotong Vocational and Technical College, Hebei Province, China
[2]Bosssoft, Fuzhou City Fujian Province, FL 350000, China

ABSTRACT

In the context of the experience economy era, museums have evolved from static exhibition spaces to multidimensional cultural experience hubs. Social media plays a crucial role in evaluating public satisfaction. This study aims to examine public perceptions of museum experiences, focusing on thematic content and overall emotional trends toward museums. While natural language processing (NLP) models have been widely applied in the field of computing, their use in museum experience design, particularly in the Chinese context, remains limited. This study employs the pre-trained model in NLP for BERTopic modeling, this research analyzes 14,800 recent online reviews (dated January 1, 2024 - December 31, 2024) from China's top 10 most-searched museums on Dianping. Generating three core topics: technological applications, cultural immersion, and service quality. Within these themes, several conflict issues emerge, such as difficulties in ticket reservation systems, the fragmentation of experiences due to digital technology, and the tension between artifact display technologies and the commercialization of cultural products. Sentiment analysis using SnowNLP indicates that while architectural design, exhibition content, and technological applications generally elicit positive emotions, satisfaction with ticketing systems and service efficiency remains low. Additionally, immersive experiences show high levels of visitor engagement, but concerns arise regarding the excessive complexity of technology, which affects the balance between technology use and historical-cultural appreciation. These findings provide valuable insights for improving museum experience design, particularly in enhancing service quality, balancing technological applications, and preserving cultural heritage.

Keywords: Museum, Bertopic model, Sentiment analysis, Exhibition experience, Social media corpus

INTRODUCTION

The sustainable development of museums requires addressing the challenges of cultural diversity within broader social transformations to meet the evolving demands of the public for cultural engagement (Wen & Ma, 2024). Traditional physical exhibitions are finding it increasingly difficult to meet the expectations of modern audiences, who seek more immersive and interactive experiences.

Research shows systematically incorporating visitor feedback into planning can markedly enhance exhibition design efficiency (Gao & Yu, 2024). Compared to traditional data - collection methods like research, interviews, and behavioral studies, there's a need to gather real - time data and conduct in - depth analysis. In comparison, BERTopic modelling has demonstrated sturdy performance in semantic analysis, especially in automation data processing and changing topic exploration. Using BERT-based systems allows for automated topic identification without requiring a specified number of themes, offering greater efficiency and processing speed when dealing with huge-level textual information compared to the LDA model (Liang et al., 2022).

This study focuses on the ten most trending museums in China in 2024, analyzing review data from the Dianping platform. By applying advanced natural language processing techniques through BERTopic modeling, the study aims to uncover key themes in visitor perceptions and analyze overall sentiment trends. Specifically, it seeks to address the following research questions:

1. What are the primary themes in public perceptions of museums?
2. What is the distribution of sentiment tendencies in public discourse on museums?

THE RELATIONSHIP BETWEEN MUSEUMS AND SOCIAL MEDIA

Current museum experience development emphasizes three interconnected trends: First, the integration of AI-driven interactive technologies facilitates immersive experiences through virtual reality (VR), augmented reality (AR), and tactile interfaces (Ioannakis et al., 2020). Second, visitor-centric service systems employ data analytics to address diverse audience needs (Di Pietro et al., 2014). Third, social media transcends physical limitations by enabling sustained knowledge exchange and cultural dialogue between museums and global audiences (Zollo et al., 2021).

As Kefi et al. (2024) demonstrate, social media enables museums to transform into inclusive, interactive spaces that foster meaningful co-creation with audiences. Specifically, museums utilize social platforms not only to enhance institutional visibility but also to curate personalized experiences through user-generated content (UGC). The real-time feedback mechanisms inherent to social media effectively capture visitors' differentiated experience data, providing actionable insights for service optimization (Gao & Yu, 2024).

RESEARCH DIRECTIONS IN MUSEUM EXPERIENCE PERCEPTION

Studies on museum experience perception primarily employ methods such as literature review, empirical research, technological innovation, and interdisciplinary approaches. The most prominent research methods include in-depth interviews and big data mining. In-depth interviews typically adopt qualitative semi-structured discussions, while big data mining focuses on extracting rich textual and visual materials from social media platforms

(Huo et al., 2024). Social media data from platforms like Instagram, Ctrip, and Dianping have been widely used in public perception research. These data not only contain abundant textual content but also include visual materials, providing valuable sources for understanding users' perceptions of cultural products (Gao & Yu, 2024). However, visual data show limitations in capturing transient scenarios, whereas textual descriptions can more comprehensively and diversely document the entire perceptual experience.

In recent years, scholars conducting textual analysis of cultural perception have increasingly focused on latent human emotions embedded in textual information. These emotions manifest through public comments, travelogues, and other content on social media platforms. Although public perceptions of cultural experiences generally remain neutral or positive, occasional negative emotions emerge, which hold instructive significance for improving experiential demands (Wang et al., 2022). However, current perception research rarely employs BERTopic modeling for thematic analysis, particularly in studies of museum experience perception within the Chinese context. BERTopic, based on pre-trained language models, demonstrates superior capability in capturing contextual relationships and semantic information between words. The bidirectional Transformer architecture of BERT enables understanding of complex semantic structures, thereby achieving more accurate topic identification (Grootendorst, 2022). Compared with traditional LDA models, BERTopic eliminates the need for predefined topic numbers, directly generating appropriate thematic clusters through its modeling process (Egger & Yu, 2022).

RESEARCH METHODOLOGY

Dianping, as a convenient and efficient social media platform, has become one of the primary channels for the public to express personal opinions and demands online (Wen & Ma, 2024). Its data, characterized by real-time responsiveness and concise expression, has been widely applied in perceptual experience studies (Zhang et al., 2018). As shown in Figure 1, the research process begins with data collection and preprocessing. Comments from the top ten most popular museums on Weibo's trending topics were crawled as analysis targets. Using web crawlers, 14,800 raw data entries were retrieved from the Dianping API between January 1, 2024, and December 31, 2024. Preprocessing involved data filtering—leveraging 35 Harbin-specific stop words and custom dictionaries—and tokenization via the Jieba tool to ensure data accuracy and usability. Subsequently, the BERTopic framework was implemented to model themes. This process integrated modules such as Transformers embeddings for contextual representation, UMAP for dimensionality reduction, HDBSCAN for clustering, Count Vectorizer for text serialization, and e-TF-IDF weighting to refine topic significance. Following topic extraction, core themes and common topics in museum experiences were identified and validated through an expert panel to ensure interpretative rigor. Finally, the preprocessed dataset underwent sentiment analysis using the Snow NLP model. This step aimed to derive overall emotional tendencies and assess the museum's emotional polarity, thereby linking thematic patterns to public sentiment.

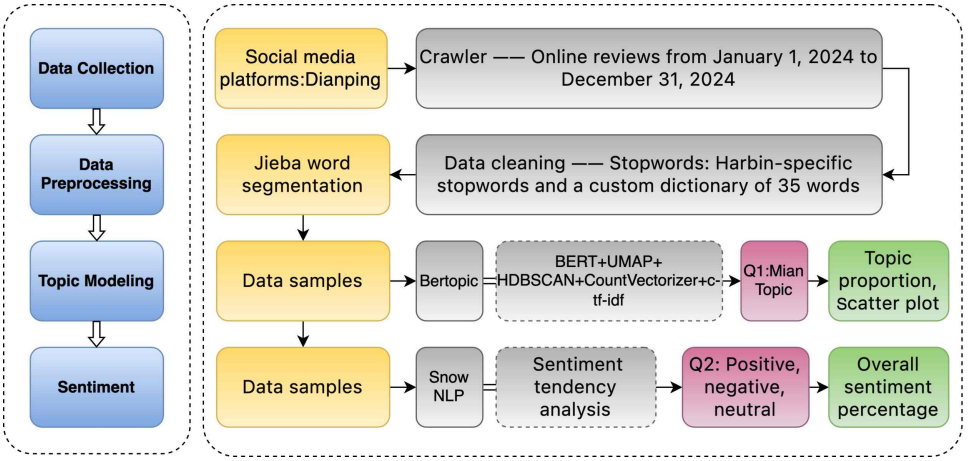

Figure 1: Flowchart of Museum research.

RESEARCH QUESTION 1: CORE THEMES IN MUSEUM EXPERIENCE PERCEPTION

In this study, the initial topic clustering yielded 36 subtopics, including a noise cluster (labeled -1). Subsequent semantic induction and expert review allowed us to consolidate these into three primary themes: service management (e.g., Topics 0, 8, 11), technological innovation (e.g., Topics 3, 4, 5), and cultural immersion (e.g., Topics 1, 2, 14), as depicted in Table 1 and Figures 2. Notably, the service management theme (particularly Topic 0) attracted the highest attention with a term frequency of 1,019, primarily addressing issues such as reservation inefficiencies in ticketing systems. In contrast, Topics 1 and 2 highlighted the richness of exhibition content, while Topic 3 emphasized the adoption of emerging technologies. The Bertopic analysis revealed a distribution of 50.8% for service management, 31.1% for technological innovation, and 18.1% for cultural immersion.

Table 1: Generated topics.

	Topic	Count	Name
Core Topic	–1	4633	1_China_comparison_hour_special
Service Management	Topic 0	1019	0_less_than_simply_scalper_system
	Topic 8	405	8_worst_thing_management_not
	Topic 10	315	10_attitude_staff_security_check_security
	Topic 11	313	11_staff_management_enter_scalper
	Topic 13	240	13_scalper_less_than_garbage_ticket_snatching
	Topic 16	154	16_management_worst_bad_Shanghai
	Topic 21	123	21_security_check_security_not_allow_storage
	Topic 25	104	25_less_than_review_elderly(public)

Continued

Table 1: Continued

	Topic	Count	Name
	Topic 27	92	27_staff_security_personnel_not_allow
	Topic 35	73	35_guide_don't_public_review_rating
Cultural Immersion	Topic 1	906	1_ Palace, Meridian Gate, Royal, Architecture
	Topic 2	847	2_ Must, Compare, Hours, Recommend
	Topic 14	233	14_ Hours, Must, Enter, Phoenix Crown
	Topic 20	125	20_ Monday, Public, Admission, Sunday
	Topic 23	118	23_ Must, Public, Remember, Compare
	Topic 24	105	24_ Refrigerator, Phoenix Crown, National Museum, Buy
	Topic 26	102	26_ Enter, National Museum, Security Check, Exit
	Topic 29	83	29_ Must, One Week, Morning, Enter
	Topic 33	75	33_ Rent, Enter, Must, Like
	Topic 34	73	34_ Republic of China, Golden Beast, Afternoon, South Museum
Technological Innovation	Topic 3	748	3_ Like, Very Large, Suitable, Special
	Topic 4	743	4_ Egypt, Ancient Egypt, Pharaoh, Mummy
	Topic 5	619	5_ Art, Every Piece, China, Experience
	Topic 6	448	6_ Understand, Convenient, Republic of China, Compare
	Topic 7	442	7_ Canal, Pavilion, Naked-eye, Projection
	Topic 9	352	9_ Architecture, I. M. Pei, Design, Modern
	Topic 12	303	12_ Transportation, Highlight, Convenient, Direct Access
	Topic 15	182	Canal_China_Cultural Heritage_Hall Number
	Topic 17	150	Scholar_Knowledge_Tour Guide_ Professional
	Topic 18	142	China_Large_National Level_First
	Topic 22	119	White Cut_Dinosaur_Jade Carving_ Fossil
	Topic 28	88	Experience_National Museum_Woman_ Ancient People
	Topic 30	82	Understand_Must_Guide_Local
	Topic 31	80	Transport_Highlight_Convenient_ Accessible
	Topic 32	76	Tianfu Square_Compare_Zhijiang_ Transport

In the Service Management and Operations (50.8%) category, Topics 0, 8, 10, and 13 represent this theme, with high-frequency keywords such as "scalpers," "ticket grabbing," and "security checks." The associated documents reveal user dissatisfaction with ticketing system flaws, notably higher crash rates than comparable venues (Topics 0, 13), cumbersome

service processes, and unprofessional staff attitudes (Topic 10). As noted by Centorrino et al. (2021), museums must balance service quality and facility comfort to meet demand, especially under high-traffic conditions. Accordingly, the current management model should adapt to these user requirements and enhance the visitor experience.

In the second major theme, Cultural Immersion and Heritage Preservation (31.1%),Topics 1, 24, and 34 revolve around keywords such as "palace," "phoenix coronet," and "history." Representative documents indicate two distinct trends in users' deep engagement with cultural heritage. On the one hand, high-value exhibits (e.g., the Phoenix Coronet of Empress Xiaoduan) elicit a strong sense of cultural identity; on the other hand, excessive commercialization (e.g., "internet celebrity check-ins" in Topic 24) fragments the viewing experience, reflecting a form of media consumption. This finding aligns with the selective spiral theory, which posits that audiences selectively filter or amplify specific information during cultural dissemination, thereby shaping their overall museum experience (Noelle-Neumann, 1974). Therefore, it is essential to strike a balance between preserving traditional culture and embracing commercialization, as well as between enhancing the viewing experience and meeting visitors' consumption demands.

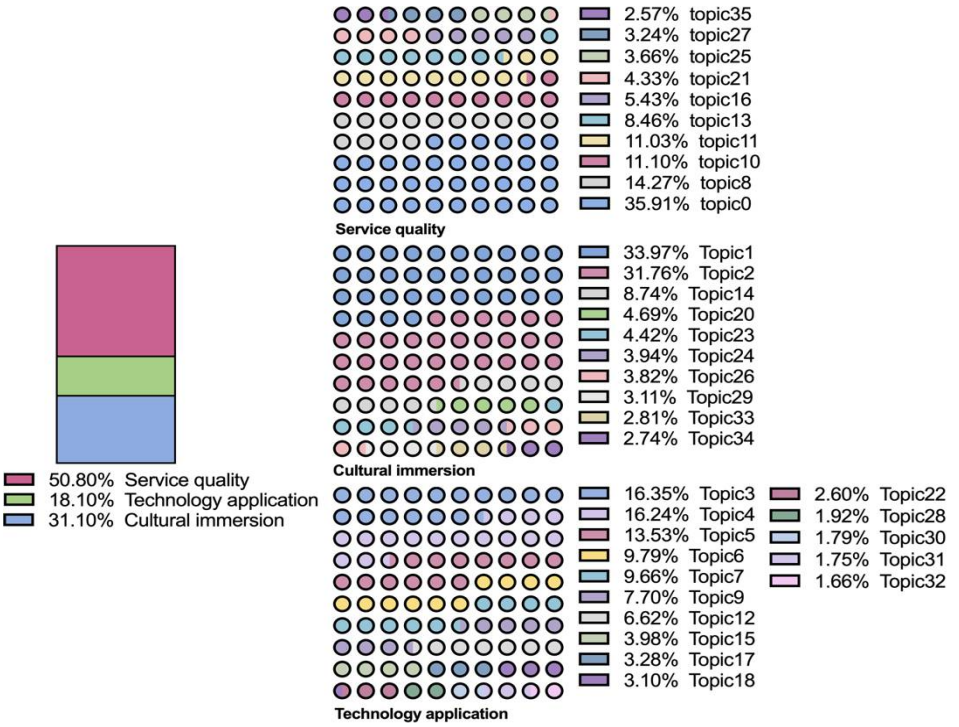

Figure 2: Proportion chart of topics and core themes.

Next, in the final Technological Applications theme, Topics 7, 15, and 22 emphasize technology-driven enhancements with keywords such as "naked-eye projection" and "digital immersion." The data indicate that museums

adopting AR/VR technologies (e.g., the Yangzhou Grand Canal Museum) can significantly boost user retention. However, audio guides described as "harsh" or "repetitive" have the potential to disrupt the visitor experience. This phenomenon supports the technological development trends highlighted by Ioannakis et al. (2020), which advocate for integrating the complete visitor experience with immersive technology applications. Moreover, it is crucial to strike a balance between leveraging technology and mitigating the adverse effects of excessive technological interference on visitor perception (Wen & Ma, 2024). In parallel, as technological innovation continues, ensuring an equilibrium between personalized, interactive customization and the museum's exhibition presentation remains essential.

As illustrated by the scatter plot (Figure 3), service quality, technology adoption, and cultural immersion exhibit strong interconnections in the dataset. Nonetheless, some points deviate from the main clusters and primarily relate to cultural immersion and heritage preservation, indicating insufficient integration of cultural creativity and user experience (blue sections, Topics 4, 14, 24). Such themes reflect visitors' unique needs and experiences regarding souvenir purchases. Although some visitors reference keywords like "queuing" and "fridge magnets," indicating partial satisfaction with cultural merchandise and exhibit presentation, issues remain—namely weak alignment between museum-branded products and institutional identity, as well as concerns about "scratches," "purchase limits," and "counterfeits."

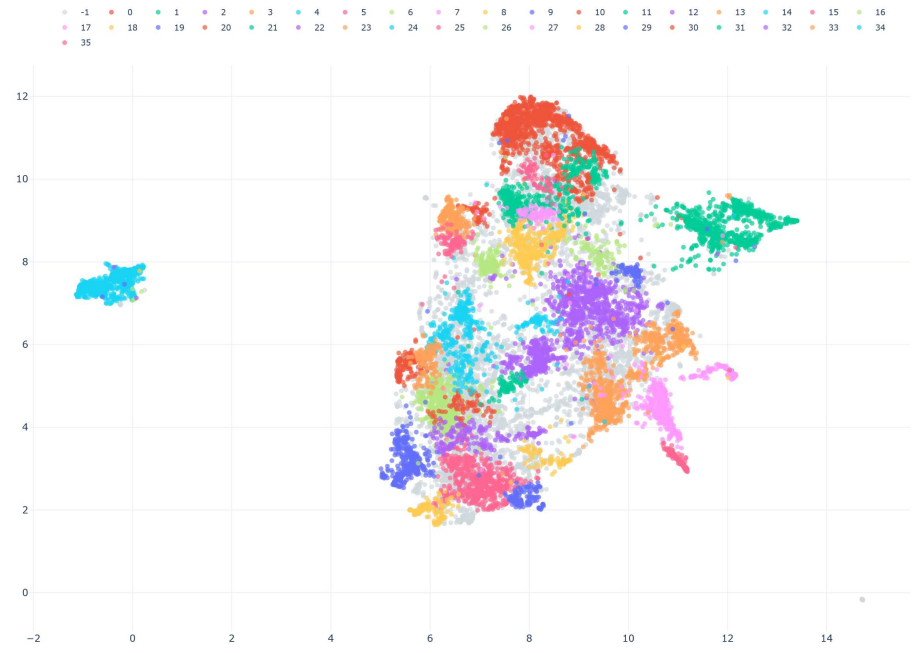

Figure 3: The visualization of scatter plots after dimensionality reduction of high-dimensional data.

RESEARCH QUESTION 2: PERCEPTUAL SEMANTIC SENTIMENT ANALYSIS OF MUSEUMS

According to the SnowNLP sentiment analysis model, as illustrated in Figure 4, visitor comments fall into three categories: positive, neutral, and negative. The results show that positive sentiment dominates at 68.1%, indicating a generally favorable perception of the Museum. Specifically, the peak density near a sentiment score of 0.0 is approximately 4, while the peak near 1.0 reaches nearly 12, suggesting that positive feedback notably exceeds neutral or slightly negative comments. Thematic analysis further reveals that most positive samples center on satisfaction with core themes such as technological applications and cultural immersion, alongside transportation accessibility, comprehensive exhibits, and the museum's architecture and technological features. Negative sentiment accounts for 29.1%, reflecting dissatisfaction with service quality in the core theme—evidenced by ticketing system failures, scalpers inflating ticket prices, and complaints about convoluted processes and inadequate service standards. Neutral sentiment comprises only 2.8%; the red curve indicates a relatively low density between the two peaks, implying that few comments fall within the 0.2 to 0.8 sentiment range. Consequently, the majority of visitor feedback exhibits a clear emotional tendency.

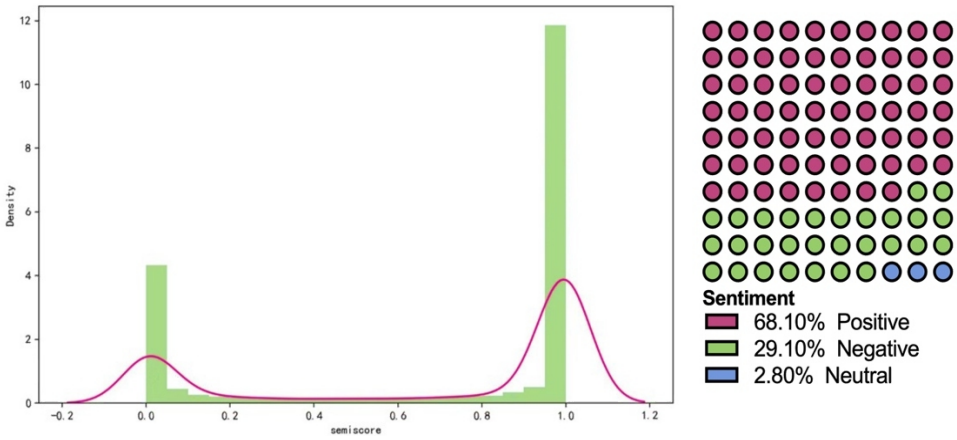

Figure 4: Proportion chart of sentiment tendencies generated by SnowNLP.

Specifically, within the positive comments, visitors consistently commend the museum's diverse range of exhibitions and professional guided services. For instance, the enigmatic allure of Sanxingdui artifacts, the dynamism of shadow puppet displays, and the exquisite craftsmanship of other historical relics all receive high praise. Scheduled complimentary tours and in-person guided explanations further enhance visitors' understanding of the exhibits' cultural and historical backgrounds. In addition, the creative product designs, when integrated with the exhibits, prove appealing, while the museum's modern architectural style and well-organized layout provide a comfortable viewing environment.

Negative sentiments primarily focus on service quality issues, including ticketing system crashes, scalpers inflating ticket prices, and the complexity of reservation and admission procedures. Excessive crowding during holidays and weekends, coupled with insufficient resting areas, also detracts from visitor comfort. Moreover, some guests feel that exhibit descriptions lack sufficient depth, and the high pricing of cultural and creative products weakens their purchasing intent. Neutral sentiments, though relatively scarce, often stem from limited visiting time or a mismatch between personal interests and exhibition themes, leading to an incomplete experience. Some visitors suggest adding more interactive elements and diversifying the exhibitions to further enrich the museum experience.

CONCLUSION AND LIMITATIONS

This study employs BERTopic topic modeling and SnowNLP sentiment analysis to conduct an in-depth exploration of social media comment data from Chinese museums. It identifies three core dimensions of the museum experience—service management, technological applications, and cultural immersion—and reveals the relationships between these themes and both visitor satisfaction and negative sentiment. The findings indicate that, in high-traffic scenarios, service management faces issues such as ticketing system failures and scalpers inflating ticket prices; cultural immersion, while fulfilling visitors' deep cultural needs, must also balance the fragmented experiences induced by commercialization; and although technological applications enhance immersion and interactivity, excessive intervention may compromise the authenticity of historical culture. Despite the predominance of positive sentiment in the comments, there remains significant room for improvement in service quality, process design, and cultural product offerings.

However, this study primarily relies on data from social media platforms such as Dianping, which may bias the sample toward more active users and raise concerns regarding representativeness. Additionally, the sentiment analysis model exhibits limitations in capturing complex linguistic phenomena, such as sarcasm and humor, which could lead to misclassification of sentiment. Future research should integrate multimodal data (e.g., images, audio, and video), conduct comparative analyses across different types and sizes of museums, and incorporate more advanced natural language processing and sentiment analysis techniques to enhance the accuracy of sentiment detection and topic extraction. These efforts will support the development of more targeted and forward-looking operational strategies for museums.

REFERENCES

Centorrino, P., Corbetta, A., Cristiani, E., & Onofri, E. (2021). Managing crowded museums: Visitors flow measurement, analysis, modeling, and optimization. Journal of Computational Science, 53, 101357. https://doi.org/10.1016/j.jocs.2021.101357

Di Pietro, L., Guglielmetti Mugion, R., Renzi, M., & Toni, M. (2014). An Audience-Centric Approach for Museums Sustainability. Sustainability, 6(9), 5745–5762. https://doi.org/10.3390/su6095745

Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. Frontiers in Sociology, 7, 886498. https://doi.org/10.3389/fsoc.2022.886498

Gao, B., & Yu, S. (2024). Upgrading museum experience: Insights into offline visitor perceptions through social media trends. Emerging Trends in Drugs, Addictions, and Health, 4, 100137. https://doi.org/10.1016/j.etdah.2023.100137

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure (arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794

Huo, H., Shen, K., Han, C., & Yang, M. (2024). Measuring the relationship between museum attributes and visitors: An application of topic model on museum online reviews. PLOS ONE, 19(7), e0304901. https://doi.org/10.1371/journal.pone.0304901

Ioannakis, G., Bampis, L., & Koutsoudis, A. (2020). Exploiting artificial intelligence for digitally enriched museum visits. Journal of Cultural Heritage, 42, 171–180. https://doi.org/10.1016/j.culher.2019.07.019

Kefi, H., Besson, E., Zhao, Y., & Farran, S. (2024). Toward museum transformation: From mediation to social media-tion and fostering omni-visit experience. Information & Management, 61(1), 103890. https://doi.org/10.1016/j.im.2023.103890

Liang, X., Hua, N., Martin, J., Dellapiana, E., Coscia, C., & Zhang, Y. (2022). Social Media as a Medium to Promote Local Perception Expression in China's World Heritage Sites. Land, 11(6). https://doi.org/10.3390/land11060841

Noelle-Neumann, E. (1974). The Spiral of Silence a Theory of Public Opinion. Journal of Communication, 24(2), 43–51. https://doi.org/10.1111/j.1460–2466.1974.tb00367.x

Wang, J., Xia, Y., & Wu, Y. (2022). Sensing Tourist Distributions and Their Sentiment Variations Using Social Media: Evidence from 5A Scenic Areas in China. ISPRS International Journal of Geo-Information, 11(9), 492. https://doi.org/10.3390/ijgi11090492

Wen, J., & Ma, B. (2024). Enhancing museum experience through deep learning and multimedia technology. Heliyon, 10(12), e32706. https://doi.org/10.1016/j.heliyon.2024.e32706

Zhang, C. X., Xiao, H., Morgan, N., & Ly, T. P. (2018). Politics of memories: Identity construction in museums. Annals of Tourism Research, 73, 116–130. https://doi.org/10.1016/j.annals.2018.09.011

Zollo, L., Rialti, R., Marrucci, A., & Ciappei, C. (2022). How do museums foster loyalty in tech-savvy visitors? The role of social media and digital experience. Current Issues in Tourism, 25(18), 2991–3008. https://doi.org/10.1080/13683500.2021.1896487