Exploring User Preferences for Museum Guides: The Role of Chatbots in Shaping Interactive Experiences

Bingqing Chen*
School of Advanced Technology
Xi'an Jiaotong-Liverpool University
Suzhou, China
bingqing.chen16@student.xjtlu.edu.cn

Ruoyu Wen*
University of Canterbury
Christchurch, New Zealand
rwe77@uclive.ac.nz

Shufang Tan
Peking University
Beijing, China
tsf@stu.pku.edu.cn

Yue Li†
School of Advanced Technology
Xi'an Jiaotong-Liverpool University
Suzhou, Jiangsu, China
yue.li@xjtlu.edu.cn

Figure 1: Illustration of interaction with three types of chatbots: (a) Docent-style chatbot using a third-person narrative; (b) Artifact chatbot using the first-person narrative of the artifact; (c) Creator chatbot using the first-person narrative of the creator.

Abstract

Museums are increasingly using chatbots to transform passive visits into interactive experiences, leveraging advancements in Large Language Models (LLMs) for more engaging interactions. However, design guidelines for chatbot roles and interactions tailored to user preferences in museum contexts remain underexplored. To address this, we conducted an online survey with 65 participants, examining preferred chatbot roles and their relationship to artifact characteristics. Participants strongly favored chatbots using a first-person narrative as artifact creators, appreciating their empathetic, immersive, and novel perspectives. The user perceptions of chatbot roles are also found to be influenced by artifact characteristics, including artifact category, its popularity, and whether it depicts human or animal figures. However, concerns about the authenticity and ethical representation of historical figures emerged. These findings provide valuable insights for designing engaging and culturally sensitive chatbot interactions in museums.

*The authors contribute equally to this work.
†Corresponding author.

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3720067

CCS Concepts

• **Human-centered computing** → *Interaction design.*

Keywords

Museum chatbots, Large Language Models (LLMs), interactive museum experiences, user preferences, chatbot roles

ACM Reference Format:
Bingqing Chen, Ruoyu Wen, Shufang Tan, and Yue Li. 2025. Exploring User Preferences for Museum Guides: The Role of Chatbots in Shaping Interactive Experiences. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3706599.3720067

1 Introduction

Museums play an increasingly vital role in society by preserving cultural heritage, promoting education [18], fostering cultural transmission, facilitating social interaction [32], contributing significantly to societal progress and economic development, and pursuing a sustainable future [11]. To meet the evolving expectations of modern audiences, museums are shifting towards more dynamic and interactive experiences. Traditionally, museum visits have been a passive experience, with visitors primarily observing exhibits and consuming information in static formats [7]. To enhance the visiting experience, museums increasingly integrate digital technologies to provide more interactive and engaging experiences [8, 28]. Chatbots systems, powered by AI and designed to facilitate personalized and interactive communication, have been widely used in museums to enhance visitor engagement and support by offering human-like conversations as virtual tour guides [2, 9, 22]. Prior research indicates that chatbots can effectively engage visitors and improve their overall museum experience [13, 19].

The advancement of large language models (LLMs) presents new opportunities to enhance visitors' interactive experiences with chatbots. Leveraging their strong natural language understanding and generation capabilities, LLMs enable chatbots to deliver more contextual and adaptive conversational interactions for users [17]. By supporting role-playing, LLM-powered chatbots can facilitate more natural, open-ended conversations, creating human-like interactions with users [4]. Nevertheless, there is a lack of user-centered guidelines for designing chatbots in the museum context, particularly regarding conversational style and interaction methods. In this study, we focus on exploring the user preferences regarding the chatbot role and design implications for interactive experiences in museums. Our guiding research questions are:

RQ1 What kind of chatbot roles do users prefer for the museum guides?
RQ2 Are there any relationships between artifact characteristics and users' perceptions of chatbot roles?

We conducted an online survey with 65 full responses, ensuring diversity in age, gender, museum visit frequency, and familiarity with generative AI. The main finding indicates a clear preference for chatbots with a first-person narrative as the artifact creator for museum guides. This approach was favored for its ability to provide a comprehensive understanding of the creation process, deemed empathetic, novel, and immersive by participants. The creator chatbot was particularly well-received for describing still-life and landscape paintings. However, the study also highlighted concerns about authenticity and ethical implications, especially in representing historical figures. These insights offer valuable guidance for designing more effective and engaging chatbot interactions in cultural heritage contexts.

2 Background And Related Work

2.1 Conversational Agents and Large Language Models

Conversational agents, or chatbots, face challenges in understanding complex contexts, generating natural dialogue, and handling open-ended queries [1, 14, 26]. Large Language Models (LLMs) have improved adaptability to complex dialogue scenarios, enhancing user experience across domains like education and entertainment. For instance, Qin et al. [23] developed CharacterMeet, a GPT-4-powered system aiding writers in character development through interactive dialogue, customizable visuals, and voice. Similarly, Zhang et al. [33] created EcoEcho, an AI-driven role-playing game where players engage with NPCs to explore sustainability challenges. In museum contexts, LLMs are used to create personalized, multilingual experiences. Trichopoulos et al. [29] integrated ChatGPT-4 and Whisper to build an immersive museum guide, while Vasic et al. [31] developed a virtual tour of the Civic Art Gallery of Ascoli, allowing users to explore artifacts interactively. Though typically used as guides, LLM-powered systems also foster critical thinking and logical reasoning, as seen in Danry et al. [5], where AI-driven question-based explanations improved users' logical discernment. As a result, this study aims to investigate how to design more engaging LLM-powered chatbots that encourage users to actively think and explore artifact-related information through interactions with role-playing AI agents. Building on the findings of Noh and Hong [19], who demonstrated that reenacted historical figures enhance user engagement and emotional connection through first-person narratives and embodied interactions, we introduce three types of AI agents for comparison: the artifact itself, the artifact's creator, and a docent-style chatbot. This approach seeks to explore which interaction style users prefer, building the foundation for future design improvements.

2.2 Intelligent Guide in Museums

A guide plays a vital role in the museum experience by conveying cultural understanding and creating meaningful interactions [6]. Research highlights that guides enhance tours by fostering positive emotions like joy and satisfaction [3]. Professional guides significantly improve visitor learning through personalized narratives, deepening connections with artifacts [6]. As noted by Origilia [20], the depth and clarity of a guide's explanation shape the richness of the visitor's experience. Intelligent virtual chatbots supported by digital technology are extensions of knowledgeable and professional tour guides, offering visitors personalized information and communication. Kiesel et al. [12] provided fundamentals for chatbot creation by exploring users' information needs by collecting user-generated questions and presenting a comprehensive database. These embodied virtual guides integrate interactive content, allowing users to explore artifacts dynamically, thereby enhancing engagement and a sense of social presence [25]. Tsitseklis et al. [30] designed a quiz-based chatbot in the form of a dialogue interface supported by natural language processing (NLP), ensuring an enriched and immersive virtual museum experience. Building on natural question-answering capabilities, researchers also endowed the chatbot with diverse roles, impacting the user experience in different ways. For instance, Saito et al. [24] found that human-like appearances enhance user satisfaction. Sylaiou et al. [27] showed that social roles (e.g., visitors, security guards, curators) affect the perceived credibility of artifact explanations. Liu et al. [15] demonstrated that advanced language model-driven personas, such as visitors, tour guides, and famous poet Li Bai, increase immersion and interest, enhancing engagement and enjoyment. Lopez et al. [10] compared two LLM-powered guide concepts, animated objects, and abstract humanoid guides, and found that participants preferred the animated objects. However, the design of chatbots in these studies often reflects designer preferences rather than user-centered or aligned with visitors' cognitive perceptions. Additionally, previous research did not study the connections between chatbot roles and the specific characteristics of artifacts, as well as their cultural background. We propose that the design of the museum chatbot should emulate the professionalism and authentic knowledge of real tour guides while imbuing them with distinct personalities to enable more engaging and immersive interactions.

Figure 2: Eight Chinese artifacts (A1-A8) and non-Chinese artifacts (A9-A16) used in the survey. Images of their corresponding creators were shown below (C1-C16).

3 Methodology

3.1 Survey Design

A virtual exhibition was created and presented using curated images to explore user preferences for three different types of chatbots in a museum setting before implementation. Sixteen artifacts and their creators were included (see Figure 2). We implemented an online survey to show the images and scripted conversations. The structure of the survey, along with the procedure, is illustrated in Figure 3. Each set of question includes a multiple-choice question ("*Please choose your most preferred chatbot guide*".) and an optional open-ended question for suggestions ("*Please provide any advice regarding your interaction, e.g., conversation with the chatbot*".) Following this, artifacts were aggregated based on the type of chatbot that participants selected, where the artifact images with the same chatbot type selected were shown on the same page. Participants were then asked to review the results, explain the reasons for their choice, and evaluate each chatbot type by filling in two established scales.

3.1.1 Selection of Artifacts and Chatbot Roles. We selected diverse artifacts from eight categories commonly displayed in museums, representing both Chinese and international creators, as shown in Figure 2 and detailed in Appendix A. The eight categories include portraits (A1, A9), animal paintings (A2, A10), landscape paintings (A3, A11), still-life paintings (A4, A12), human statues (A5, A13), animal statues (A6, A14), architecture (A7, A15), and decorative items (A8, A16).

3.1.2 Chatbot Role and Interaction Design. The information fed into the GPT-4o[1] is sourced from official museum databases and websites to ensure accuracy and authenticity. The chatbot prompts include a structured content format that provides an introduction to the artifact, details about its current location, a description of the creator, and creation background. Conversations are designed to be concise, limiting each question and answer to approximately 30 words while maintaining a lively and engaging tone through a five-round dialogue. To ensure reenacted and unique interactions while reducing redundancy in content delivery, we also prompt ChatGPT with "Distinct personas for each creator and artifact based on the information provided". Three chatbot roles with features are demonstrated in Table 1. The panel design for chatting with the chatbot is based on a typical chat dialogue box, as illustrated in Figure 1. The `DC` uses a default avatar image to reduce bias. The `AC`'s avatar is the artifact itself (A1-A16), and the `CC`'s avatar is the

[1] https://openai.com/index/hello-gpt-4o/

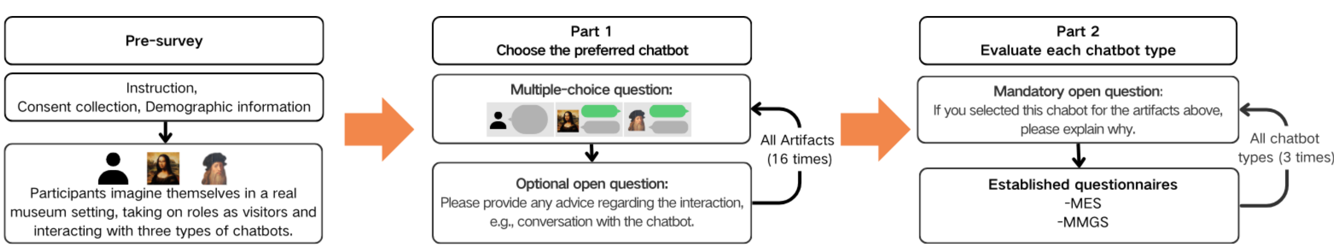

Figure 3: A flow chart illustrating the structure of the survey and the procedure.

Table 1: Features of three types of chatbots.

Role	Description	Narrator style	Language style
Docent-style chatbot (DC)	A docent-style guide that mimics guidance in physical museums.	Third-person	Phrase
Artifact chatbot (AC)	An artifact chatbot with an anthropomorphic personality.	First-person	Personification
Creator chatbot (CC)	A creator chatbot that adopts the persona of the creator of the artifact.	First-person	Reenacted

creator's image (C1-C16). For the anonymous creator cases, they were also assigned the default avatar image, ensuring consistency in presentation and maintaining a cohesive user experience. All images are cropped to a 1:1 ratio.

3.2 Measures

The Museum Experience Scale (MES) [21] is used to evaluate the influence of multimedia guides on user experiences within museums across four components: engagement, knowledge/learning, meaningful experience, and emotional connection (1 "strongly disagree" to 5 "strongly agree"). Similarly, the Multimedia Guide Scale (MMGS) [21] assesses the usefulness and usability of multimedia guides in museums, focusing on general usability, learnability and control, and quality of interaction.

3.3 Procedure

The procedure and structure of the survey are illustrated in Figure 3. Surveys were distributed online and promoted through social media for remote completion. We first collected participants' demographic information, museum visit frequency, and familiarity with generative AI. As a preliminary study for chatbot preference, participants did not chat with different types of chatbots, but were asked to read and choose between different scripted conversations. Initially, participants were provided with four textual options. After selecting an option, a corresponding scripted interaction example appears. The order of the option and question were both counterbalanced. Participants were also prompted with an optional open-ended question for advice regarding the interaction (e.g., specific feedback regarding the interaction with the certain artifact). The suggestions served as a supplement to the subsequent mandatory open questions in the second part and were summarized into possible design options, which will be carefully considered in future work. In the second part, participants completed the MES and MMGS for all three chatbot types, along with an open-ended question explaining their chatbot preferences.

3.4 Participants

A total of 133 questionnaires were started, with 68 of them remaining incomplete. This is likely attributed to the presence of repetitive questions and several open-ended items [16]. The final sample consisted of 65 participants (37 females, 28 males) aged 19 to 50 years old ($M = 26.73$, $SD = 5.74$). More than half ($N = 28$) visited museums once a year or less. Participants reported moderate familiarity with generative AI technology, with a mean score of $M = 3.22$ ($SD = 1.08$) on a scale from 1 (not familiar at all) to 5.

4 Results

4.1 Data Processing

For quantitative analysis, we used the Shapiro-Wilk test to assess data distribution, revealing non-normal distribution across all datasets. We then utilized Wilcoxon signed-rank tests to compare seven dimensions across three chatbot types. For qualitative analysis, two researchers clustered all collected feedback into pros and cons. Results are summarized in Table 2. Specifically, positive feedback rate refers to the percentage of received feedback that is expressed positively versus the all comments collected ($n = 65$).

4.2 Quantitative Analysis

4.2.1 Multiple-choice questions. In the first part of the survey, participants selected their preferred chatbot interaction across 1,040 responses (65 participants × 16 questions). The CC was selected most frequently ($n = 384$), preferred for landscape and still-life paintings (e.g., A3: 50.8%, A4: 55.4%, A11: 58.5%, A12: 66.2%). The AC ranked the second on frequency ($n = 346$), mainly selected for artifacts featuring living beings (e.g., A6: 52.3%, A9: 50.8%, A13: 58.5%). The DC was selected only 295 times, mainly for architecture (e.g., A7: 40%, A14: 47.7%). Fifteen participants selected "other", with three of whom did not provide specific suggestions. The remaining responses are summarized in section 4.3.4.

4.2.2 MES. Results for MES are shown in Figure 4(a-d). For the *engagement* subscale, a Friedman test revealed a significant difference between the three guide methods ($\chi^2(2) = 46.507, p < 0.001$) with a moderate effect size ($W = 0.358$). The following pairwise Wilcoxon signed-rank tests showed higher engagement for CC ($Z = -4.692, p < 0.001$) and AC ($Z = -5.340, p < 0.001$) compared to DC. For the *knowledge/learning* subscale, a small effect size was found ($W = 0.129$), and significant differences were revealed between AC and DC ($Z = -2.023, p = 0.043$), CC and DC ($Z = -3.890, p < 0.001$), and CC and AC ($Z = -2.324, p = 0.020$). The *meaningful experiences* subscale also showed significant differences ($\chi^2(2) = 24.311, p < 0.001$) and a small effect size ($W = 0.187$), with CC and AC outperforming DC, and CC rated higher than AC ($Z = -4.710, p = 0.018$). Finally, the *emotional connection* subscale showed significant effects ($\chi^2(2) = 32.813, p < 0.001$) with a small effect size ($W = 0.252$), and both AC ($Z = -4.463, p < 0.001$) and CC ($Z = -5.402, p < 0.001$) eliciting stronger emotional connections than DC.

4.2.3 MMGS. Results for MMGS are illustrated in Figure 4(e-g). Significant differences were found in the *learnability and control* and *quality of interaction* subscales. For *learnability and control* showed a significant difference ($\chi^2(2) = 13.549, p = 0.001$) with a small effect size ($W = 0.104$). Post-hoc tests showed the CC scored higher than the DC ($Z = -3.043, p = 0.002$) and AC ($Z = -2.979, p = 0.003$).

Table 2: Summary of three types of chatbots for positive feedback rate, Frequency of choice, matching artifacts, pros and cons.

	Docent-style Chatbot (DC)	Artifact Chatbot (AC)	Creator Chatbot (CC)
Positive feedback rate (%)	70.8%	78.5%	**84.6%**
Frequency of choice (n)	295	346	**384**
Well-suited artifact	Architecture & Unfamiliar artifacts	Portrait & Animal	Landscape & Still-life
Pros (n)	Objective (11)	Immersive (12)	Enhance understanding of artifact meaning and creation process (32)
	Comprehensive (10)	Emotional engagement (10)	Empathetic (17)
	Easy to access (2)	Easy to understand (8)	Immersion (4)
	Comfortable (2)	Sense of vitality (4)	Novelty (2)
		Novelty (4)	
Cons (n)	Too traditional (5)	Lacks souls (4)	Too subjective (3)
	Lack immersion (5)	Fail to convey insights (4)	Not convincing (3)
	Lack customization (2)		
	Too lengthy (1)		

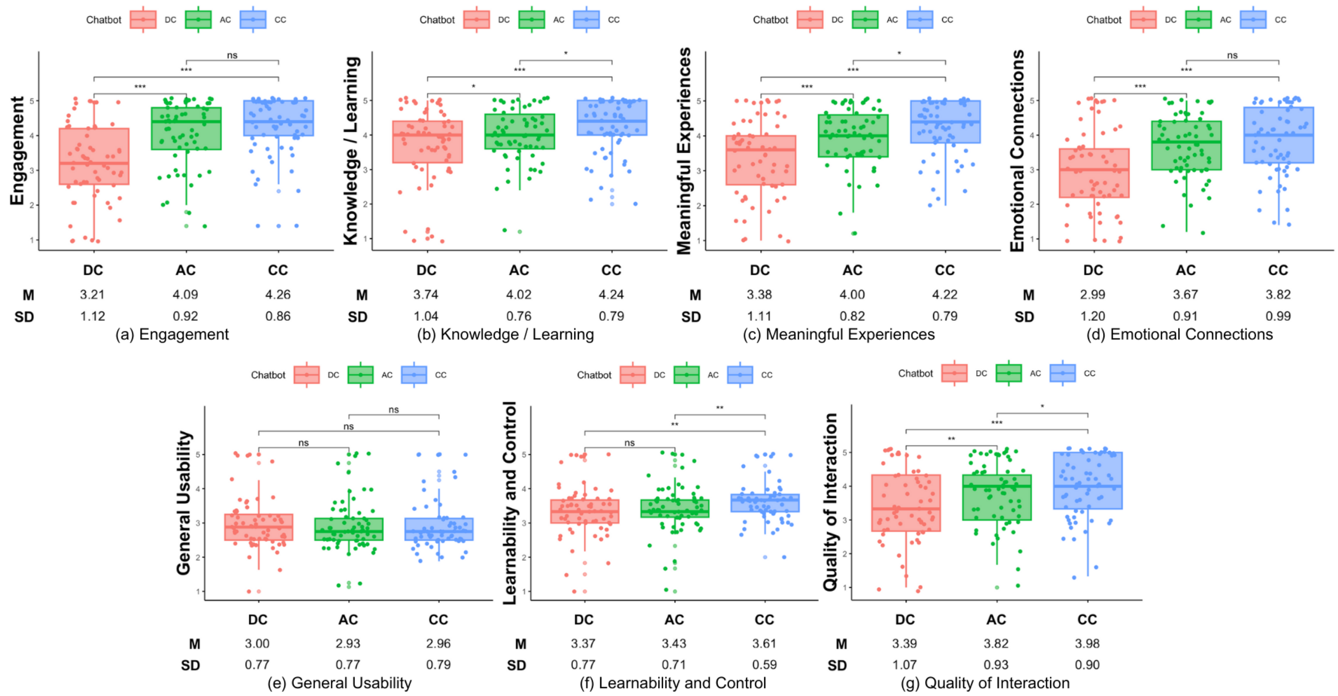

Figure 4: Box-plots and tables of descriptive statistics (means and standard deviations) showing the data analysis results of the MES and MMGS questionnaire. *DC*: docent-style chatbot. *AC*: artifact chatbot. *CC*: creator chatbot. Significance p-value: $^{*}: p < 0.05$, $^{**}: p < 0.01$, $^{***}: p < 0.001$, ns: not significant

Similarly, for *quality of interaction* subscale, significant differences ($\chi^2(2) = 12.446, p = 0.002$) and a small effect size ($W = 0.096$) were observed, with the CC rated significantly higher than the DC ($Z = -3.990, p < 0.001$) and AC ($Z = -2.397, p = 0.017$).

4.3 Qualitative Analysis

4.3.1 Creator Chatbot. Participants gave the most positive feedback on the CC, aligned with the quantitative results. About 84.6% of participants (55/65) provided positive feedback, with nearly half (32/65) noting that this approach helped them better understand the artifact's meaning and creative process to varying degrees, particularly for abstract works like landscape and still-life paintings. P11 said, *"Most of these works incorporate the author's own reflections. Engaging in a direct conversation with the author allows for deeper insight into their thoughts during the creation process, which can inspire further contemplation"*. Four participants specifically expressed a strong desire to interact with a human-form avatar since conversations with the creator provide a strong sense of connection. As P5 stated, *"objects without life struggle to empathize with thoughts and emotions"*. Additionally, the novelty (2/65), immersion (4/65), and empathy provided (17/65) of CC compared to the others were highlighted.

4.3.2 Artifact Chatbot. Approximately 78.5% of participants (51/65) provided positive feedback on the novel interaction with the anthropomorphic AC. From the perspective of interaction experience, participants found this method more immersive (12/65). As P15 said, *"It brought the artifact closer to the viewer (P15)"*. Also, this type of chatbot could provide more emotional engagement (10/65) and make the content easier to understand (8/65). Regarding the type of artifact, some participants (14/65) preferred engaging in dialogue with specific living elements (e.g., human-like or animal figures) within the artifact rather than the entire piece. As P65 emphasized, these elements allowed them to *"feel a strong sense of vitality (P65)"*. This approach also brought a novel perspective to participants. Eight participants agreed that AC makes the artifact feel more dynamic while stimulating curiosity and exploration. However, some participants expressed that, as a medium created by the artist, the artifact lacks a soul and cannot authentically convey the artist's thoughts (4/65).

4.3.3 Docent Chatbot. The DC received the lowest positive evaluation, with 70.8% (46/65). Notable drawbacks include being too traditional (5/65), lacking immersion (5/65), being too lengthy (1/65), and lacking customization (2/65). For instance, P2 expressed a desire to explore the artist's creative journey and the artifact's intricacies, while DC failed to provide insights. On the other hand, some participants appreciated the DC's objectivity (11/65), comprehensiveness (10/65), ease of access (2/65), and comfortable (2/65). For artifacts without lives, such as architecture, participants (10/65) preferred the docent-style narration as they found it clear, concise, and more suitable for conveying simple information and atmosphere. As P33 and P35 indicated, these objects themselves *"did not seem to have the ability to speak (P33, P35)"*. Another category deemed suitable for DC comprises unfamiliar artifacts and creators, as noted by 10 of the 65 participants. As P11 mentioned, he lacked interest in asking questions for artifacts that he barely knows and found the brief overview ahead more appropriate. Furthermore, four participants emphasized that in museums, it is important to enjoy the artifacts in silence rather than engage in conversation.

4.3.4 Other Suggestions for Interaction. Participants preferred conversing with the former wearer for wearable decorative items, such as the crown (A16), instead of the provided chatbot types (2/65). As P12 stated, *"Using characters connected to the artwork to narrate the story feels more engaging"*. For unfamiliar artifacts, a subset of participants (10/65) suggested combining docent-style narration (DC) and creator chatbots (CC), noting that DC could provide foundational context while CC could offer deeper insights into the artifact's significance. Seven participants suggested representing the chatbot as a 3D avatar, supporting voice interaction, and performing actions during conversations. Some of them (4/65) also preferred that information about specific artifact details be directly linked to those details, such as highlighting relevant parts during discussions. Also, three participants emphasized incorporating the artifact's historical context to enrich interactions.

5 Discussions

5.1 User Preferences of Museum Guide Chatbots

Our research illustrates that the creator chatbot (CC) emerged as the participants' most preferred option as a museum guide (RQ1). The results indicate that CC has a notable advantage over DC and AC across six key dimensions regarding user experience and usability. These results emphasize the effectiveness of integrating reenacted and first-perspective CC chatbots in creating immersive and engaging museum experiences. Also, all three types of museum guide chatbots received positive feedback, with at least 70% of responses being favorable. This indicates that visitors widely appreciate the chatbot's ability to engage, inform, and facilitate interactions, regardless of its specific role in the museum setting.

5.2 User Perception and Artifact Characteristics

This study reveals that user perceptions of chatbot roles are significantly influenced by artifact characteristics, including artifact popularity, artifact category, and whether the artifact depicts human or animal figures (RQ2). Specifically, users tended to have more positive interactions with AC or CC when the artwork was well-known, suggesting that familiarity with the piece enhanced their engagement with the chatbot. Additionally, artifacts featuring human representations were often associated with more personalized and empathetic chatbot responses, as users felt a stronger connection to the characters portrayed. These findings suggest that the characteristics of the artifact itself (e.g., artifact category, popularity, and human/animal figures) play a significant role in shaping how users perceive and interact with LLM-driven chatbots in museum settings.

5.3 Limitations and Future Work

The current study adopted a user-centered approach to explore preferred chatbot types, collect user expectations, and examine how user perceptions relate to artifact characteristics. However, the persona of each chatbot was created based on prompting, which may limit the chatbot's ability to provide fully comprehensive and accurate answers. Additionally, mitigating biased or hallucinated responses, especially when addressing sensitive topics like contemporary political issues, remains a challenge. Besides, the experiment was conducted in the form of a survey, where participants selected pre-generated dialogue examples instead of having real-time interaction with chatbots. While this ensured consistent content, it also restricted interaction opportunities. Future work implementing interactable chatbots will yield increased ecological validity. Moreover, using only a default avatar for anonymous creators may limit engagement. Allowing avatar selection or personalization could enhance the experience.

We also aim to enhance creator chatbots by fine-tuning LLM, like GPT-4o, to create more detailed, authentic, and interactable roles. We also plan to use prompt engineering to define the chatbot's conversational scope. When users input beyond this scope, the chatbot could gently steer the conversation back to the artifact.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (62207022).

References

[1] Ahmad Abdellatif, Diego Costa, Khaled Badran, Rabe Abdalkareem, and Emad Shihab. 2020. Challenges in Chatbot Development: A Study of Stack Overflow Posts. In *Proceedings of the 17th International Conference on Mining Software Repositories* (Seoul, Republic of Korea) *(MSR '20)*. Association for Computing Machinery, New York, NY, USA, 174–185. doi:10.1145/3379597.3387472

[2] Bayan AbuShawar and Eric Atwell. 2015. ALICE chatbot: Trials and outputs. *Computación y sistemas* 19, 4 (2015), 625–632.

[3] Alex Altieri, Silvia Ceccacci, Luca Giraldi, Alma Leopardi, Maura Mengoni, and Abudukaiyoumu Talipu. 2021. Affective Guide for Museum: A System to Suggest Museum Paths Based on Visitors' Emotions. In *Universal Access in Human-Computer Interaction. Design Methods and User Experience*. Springer International Publishing, Cham, 521–532.

[4] Nuo Chen, Yan Wang, Yang Deng, and Jia Li. 2025. The Oscars of AI Theater: A Survey on Role-Playing with Language Models. arXiv:2407.11484 [cs.AI] https://arxiv.org/abs/2407.11484

[5] Valdemar Danry, Pat Pataranutaporn, Yaoli Mao, and Pattie Maes. 2023. Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy over Causal AI explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 352, 13 pages. doi:10.1145/3544548.3580672

[6] Marko D. Petrović Dunja Demirović Bajrami, Nikola Vuksanović and Tatiana N. Tretiakova. 2022. Competencies of a Museum Guide as Predictors of Visitors' Learning Outcomes: A Case from Canada. *Journal of Museum Education* 47, 2 (2022), 251–262. doi:10.1080/10598650.2022.2062542 arXiv:https://doi.org/10.1080/10598650.2022.2062542

[7] Khloud Eldamshiry and Mohamed Khalil. 2018. Museum Visitors Learning Identities Interrelationships with Their Experiences. In *Green Heritage Conference: Chance - Change - Challenge*. Elain Publishing Company, The British University in Egypt.

[8] Islam Elgammal, Marco Ferretti, Marcello Risitano, and Annarita Sorrentino. 2020. Does digital technology improve the visitor experience? A comparative study in the museum context. *International Journal of Tourism Policy* 10, 1 (2020), 47–67.

[9] Giuliano Gaia, Stefania Boiano, and Ann Borda. 2019. *Engaging Museum Visitors with AI: The Case of Chatbots.* Springer International Publishing, Cham, 309–329. doi:10.1007/978-3-319-97457-6_15

[10] Irene Lopez Garcia, Ephraim Schott, Marcel Gohsen, Volker Bernhard, Benno Stein, and Bernd Froehlich. 2024. Speaking with Objects: Conversational Agents' Embodiment in Virtual Museums. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE Computer Society, Bellevue, WA, USA, 279–288. doi:10.1109/ISMAR62088.2024.00042

[11] Brown Karen. 2019. Museums and Local Development: An Introduction to Museums, Sustainability and Well-being. *Journal of Museum International* 71 (2019), 1–13.

[12] Johannes Kiesel, Volker Bernhard, Marcel Gohsen, Josef Roth, and Benno Stein. 2022. What is That? Crowdsourcing Questions to a Virtual Exhibition. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 358–362. doi:10.1145/3498366.3505836

[13] Stefan Kopp, Lars Gesellensetter, Nicole C Krämer, and Ipke Wachsmuth. 2005. A conversational agent as museum guide–design and evaluation of a real-world application. In *Intelligent Virtual Agents: 5th International Working Conference, IVA 2005, Kos, Greece, September 12-14, 2005. Proceedings 5*. Springer, Kos, Greece, 329–343.

[14] Chien-Chang Lin, Anna YQ Huang, and Stephen JH Yang. 2023. A review of AI-driven conversational chatbots implementation methodologies and challenges (1999–2022). *Sustainability* 15, 5 (2023), 4012.

[15] Dong Liu, NengLun Chen, Xu Lang, ZhiGeng Pan, Hongyi Ren, SiYuan Lin, MingMin Zhang, HuaBiao Li, and QianWei Huang. 2025. Exploring distilled spirits brewing: Utilizing multimodal interaction and intelligent virtual avatars in a VR liquor culture museum. *Entertainment Computing* 52 (2025), 100909. doi:10.1016/j.entcom.2024.100909

[16] Mingnan Liu and Laura Wronski. 2018. Examining Completion Rates in Web Surveys via Over 25,000 Real-World Surveys. *Social Science Computer Review* 36, 1 (2018), 116–124. doi:10.1177/0894439317695581 arXiv:https://doi.org/10.1177/0894439317695581

[17] Bonan Min, Hayley Ross, Elior Sulem, Amir Pouran Ben Veyseh, Thien Huu Nguyen, Oscar Sainz, Eneko Agirre, Ilana Heintz, and Dan Roth. 2023. Recent advances in natural language processing via large pre-trained language models: A survey. *Comput. Surveys* 56, 2 (2023), 1–40.

[18] Monika Murzyn-Kupisz and Jarosław Działek. 2013. Cultural heritage in building and enhancing social capital. *Journal of Cultural Heritage Management and Sustainable Development* 3 (05 2013), 35–54. doi:10.1108/20441261311317392

[19] Yeo-Gyeong Noh and Jin-Hyuk Hong. 2021. Designing reenacted chatbots to enhance museum experience. *Applied Sciences* 11, 16 (2021), 7420.

[20] Antonio Origlia, Renata Savy, Violetta Cataldo, Loredana Schettino, Alessandro Ansani, Isora Sessa, Alessandra Chiera, and Isabella Poggi. 2019. Human, All Too Human: Towards a Disfluent Virtual Tourist Guide. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization* (Larnaca, Cyprus) *(UMAP'19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 393–399. doi:10.1145/3314183.3323866

[21] Mohd Kamal Othman, Helen Petrie, and Christopher Power. 2011. Engaging Visitors in Museums with Technology: Scales for the Measurement of Visitor and Multimedia Guide Experience. In *Human-Computer Interaction – INTERACT 2011*. Vol. 6949. Springer Berlin Heidelberg, Berlin, Heidelberg, 92–99. doi:10.1007/978-3-642-23768-3_8

[22] Pat Pataranutaporn, Valdemar Danry, Joanne Leong, Parinya Punpongsanon, Dan Novy, Pattie Maes, and Misha Sra. 2021. AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence* 3 (12 2021). doi:10.1038/s42256-021-00417-9

[23] Hua Xuan Qin, Shan Jin, Ze Gao, Mingming Fan, and Pan Hui. 2024. CharacterMeet: Supporting Creative Writers' Entire Story Character Construction Processes Through Conversation with LLM-Powered Chatbot Avatars. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1051, 19 pages. doi:10.1145/3613904.3642105

[24] Mari Saito. 2023. Effects of Presentation Modalities in Virtual Museum Guides on Agent Impressions and Painting Evaluations.. In *Proceedings of the 11th International Conference on Human-Agent Interaction* (Gothenburg, Sweden) *(HAI '23)*. Association for Computing Machinery, New York, NY, USA, 446–448. doi:10.1145/3623809.3623958

[25] Susanne Schmidt, Gerd Bruder, and Frank Steinicke. 2019. Effects of virtual agent and object representation on experiencing exhibited artifacts. *Computers & Graphics* 83 (2019), 1–10. doi:10.1016/j.cag.2019.06.002

[26] Heung-Yeung Shum, Xiao-dong He, and Di Li. 2018. From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering* 19 (2018), 10–26.

[27] Stella Sylaiou, Vlasios Kasapakis, Elena Dzardanova, and Damianos Gavalas. 2019. Assessment of Virtual Guides' Credibility in Virtual Museum Environments. In *Augmented Reality, Virtual Reality, and Computer Graphics*, Lucio Tommaso De Paolis and Patrick Bourdot (Eds.). Springer International Publishing, Cham, 230–238.

[28] Francesca Taormina and Sara Bonini Baraldi. 2022. Museums and digital technology: a literature review on organizational issues. *European Planning Studies* 30, 9 (2022), 1676–1694. doi:10.1080/09654313.2021.2023110 arXiv:https://doi.org/10.1080/09654313.2021.2023110

[29] Georgios Trichopoulos, Markos Konstantakis, George Caridakis, Akrivi Katifori, and Myrto Koukouli. 2023. Crafting a Museum Guide Using ChatGPT4. *Big Data and Cognitive Computing* 7, 3 (2023), 148.

[30] Konstantinos Tsitseklis, Georgia Stavropoulou, Anastasios Zafeiropoulos, Athina Thanou, and Symeon Papavassiliou. 2023. RECBOT: Virtual Museum navigation through a Chatbot assistant and personalized Recommendations. In *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization* (Limassol, Cyprus) *(UMAP '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 388–396. doi:10.1145/3563359.3596661

[31] Iva Vasic, Hans-Georg Fill, Ramona Quattrini, and Roberto Pierdicca. 2024. LLM-Aided Museum Guide: Personalized Tours Based on User Preferences. In *Extended Reality*, Lucio Tommaso De Paolis, Pasquale Arpaia, and Marco Sacco (Eds.). Springer Nature Switzerland, Cham, 249–262.

[32] Mao-Ying Wu, Yixuan Tong, Geoffrey Wall, and Tianyu Ying. 2021. Cultural production and transmission in museums: A social practice perspective. *Annals of Tourism Research* 87 (2021), 103130. doi:10.1016/j.annals.2020.103130

[33] Qinshi Zhang, Ruoyu Wen, Latisha Besariani Hendra, Zijian Ding, and Ray LC. 2025. Can AI Prompt Humans? Multimodal Agents Prompt Players' Game Actions and Show Consequences to Raise Sustainability Awareness. arXiv:2409.08486 [cs.HC] https://arxiv.org/abs/2409.08486

Appendix A: Details about the artifacts and creators used in the survey

Artifact No.	Name	Creator No.	Creator Name
$A1$	Court Ladies Adorning Their Hair with Flowers	$C1$	Anonymous
$A2$	White Goose and Red Polygonum	$C2$	Mengfu Zhao
$A3$	Mountain Dwelling on a Summer Day	$C3$	Meng Wang
$A4$	Basket of Flowers	$C4$	Anonymous
$A5$	Judge of Hell	$C5$	Anonymous
$A6$	Pottery Xiezhi-unicorn	$C6$	Anonymous
$A7$	Shuangta	$C7$	Wenhan Wang
$A8$	Chinese Antique Yuan Meiping	$C8$	Anonymous
$A9$	The Mona Lisa	$C9$	Leonardo Da Vinci
$A10$	The Monarch of the Glen	$C10$	Edwin Landseer
$A11$	Impression, Sunrise	$C11$	Claude Monet
$A12$	Sunflowers	$C12$	Vincent van Gogh
$A13$	David	$C13$	Michelangelo
$A14$	Horse of Selene	$C14$	Anonymous
$A15$	Florence Cathedral	$C15$	Filippo Brunelleschi
$A16$	Holy Crown of Hungary	$C16$	Anonymous

Appendix B: Detailed demographic information of participants

Category	Group	Sample (n)	Percent (%)
Gender	Male	28	43.08
	Female	37	56.92
Age	18-20	6	9.23
	21-30	47	72.31
	31-40	11	16.92
	40+	1	1.54
Frequency of museum visiting	Very frequently (once a month or more)	2	3.08
	Frequently (a couple of times a year)	16	24.62
	Occasionally (2-3 times a year)	19	29.23
	Rarely (once a year or less)	28	43.08
	Never	0	0.00
Familiarity with Generative AI	1 (not familiar at all)	4	6.15
	2	12	18.46
	3	23	35.38
	4	18	27.69
	5 (very familiar)	8	12.31