Digital Applications in Archaeology and Cultural Heritage 42 (2026) e00585

Contents lists available at ScienceDirect

Digital Applications in Archaeology and Cultural Heritage

journal homepage: www.elsevier.com/locate/daach

Chatty art labels: Evaluation of a conversational museum guide

Tobias Tue, Henning Pohl, Timothy Merritt *

Aalborg University, Selma Lagerløfs Vej 300, Aalborg, 9220, Denmark

ARTICLE INFO

Keywords:
Large language models
Museum guide
Augmented reality

ABSTRACT

This study explores the potential of a large language model (LLM)-powered Conversational Agent (CA) as a virtual museum guide to enhance visitor experi-ences. A prototype was developed, enabling natural language interactions with an LLM-powered CA about augmented reality (AR) artwork viewed through a head-mounted device (HMD). An exploratory field study with 22 museum visitors found positive evaluations based on a multimedia guide assessment scale. Results showed that the most prevalent were fact-based questions, while qualitative findings highlighted appreciation for personalized information, concerns over information gaps, challenges in formulating questions, and discomfort with public voice interactions. The study underscores the potential for LLM-powered guides for museum experiences, while identifying design challenges for future research and development.

1. Introduction

Museum visitors need and desire 'free-choice' to engage with exhibits and information as they see fit, often in a non-linear manner (Falk and Dierking, 2018). Museums, therefore, strive to provide a range of options for visitors, such booklets, audio guides, labels next to individual pieces, museum apps, and guided tours. Booth (1998) has described this as the *information needs of visitors* and pointed out that what is needed differs between visitor groups. However, much of this information is mostly static or a collection of linear content that is the same for all visitors. Formats that can be highly personalized, such as guided tours, generally scale poorly and might require substantial expertise to deliver. Guided tours often cater to group requirements, which then take priority over individual preferences. Consequently, making museums more interactive and adaptive for the individual visitor can be, as Stock et al. (2007) pointed out, an *"significant investment"*.

With advances in large language models (LLMs), such as fine-tuning (Wei et al., 2024) or retrieval augmented generation (Banerjee et al., 2024) that has shown utility for supporting tourist guidance, conversational museum guides are becoming a feasible alternative to create more adaptive and interactive experiences (Trichopoulos et al., 2023; Trichopoulos, 2023; Vasic et al., 2024). Such models are able to process natural language user queries and can also respond in kind (Schaffer et al., 2023, 2024; Dugulean et al., 2020; Gaia et al., 2019), acting sometimes as a form of virtual tour guide (Vasic et al., 2024; Maniscalco et al., 2024; Tokuno et al., 2024). Some multimodal models are also able to directly recognize objects and contexts relevant to a query from a picture, video, or audio recording made by the visitor (Lee et al., 2024). Furthermore, LLMs can be given knowledge banks or access to external information sources, in order to guide and constrain the answers they provide. As Bin et al. (2024) have demonstrated, LLMs can also be further tuned for painting analysis and description.

2. Research aim

While the technical possibilities of LLMs in museums are substantial, recent studies have begun to examine their implications for museum experiences. Chen et al. (2025) investigated preferences for different museumchatbot roles, highlighting interest in first-person creator narratives while also identifying concerns regarding authenticity and ethical representation. Wang and Deterding (2026) proposed a conceptual framework for understanding conversational agents as technologies that mediate visitors' relationships with cultural heritage. Lin and Hu (2026) developed and evaluated MuseForge, demonstrating how generative AI can support engagement, motivation, and creative learning in digital museum education. Together, these studies establish an emerging foundation for understanding the design and experience of generative-AI systems in cultural heritage contexts. However, empirical knowledge remains limited regarding how visitors engage with voice-based LLM guides during situated museum visits, including how they formulate questions, select and evaluate information, and negotiate the social and interactional demands of speaking with an AI guide in a

* Corresponding author.
E-mail addresses: ttue22@student.aau.dk (T. Tue), henning@cs.aau.dk (H. Pohl), merritt@cs.aau.dk (T. Merritt).

https://doi.org/10.1016/j.daach.2026.e00585
Received 13 October 2025; Received in revised form 13 June 2026; Accepted 4 August 2026
Available online 13 August 2026
2212-0548/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

public setting. To investigate these issues, we implemented a conversational museum guide and evaluated it through a field study. Based on interviews and analyses of visitors' interactions with the guide, we derive implications for future research and identify important next steps for the design of conversational museum experiences.

3. Materials and methods

We implemented a mixed reality prototype of a conversational museum guide and then used it for a field study in a museum lobby.

3.1. Prototype

Our mixed reality conversational museum guide uses virtual art pieces for testing, which (1) provides more freedom and control when picking them, and (2) removes the requirement for high-quality tracking and recognition of the artwork. Furthermore, mixed reality gave participants consistency of space (retaining the museum context where the study was situated) and allowed them to see the experimenter.

We developed the prototype in Unity V2022.3.20f1 and Meta's OpenXR SDK. With passthrough mode enabled and running on a *Meta Quest 2* HMD, this allowed for the mixing of virtual world and real space. To support conversational interactions, we utilized the *NPC AI Engine* from ConvAI[1] in our project. ConvAI was configured to use the GPT-4o model using the API. This model was selected based on the compatible model suggested by the vendor at the time. Informal pilot testing was conducted within our research lab to ensure a reasonable response time that supports conversational interactions and maintains context and relevance to the artwork's topic.

For the artwork, we chose four pieces from the 19th and 20th centuries (see Fig. 1). These cover a range of styles and are generally well-known and there are extensive descriptions and analyses of them available online. Conversation calls through Convai utilized a custom prompt and a text file for information about each painting. We copied the informational text from online sources[2] and wikipedia articles[3] for each painting so that the LLM calls would have sufficient information to build a detailed response without the delay of querying external sources.

Within the constraints of the Convai tool, we created a character for each of the paintings and in the "Character Description" console defined the settings for the response style and length in the "Core Description" setting. This sets the context for any API call to the character. When API calls are made, the system uses retrieval augmented generation pulling in documents in the "Knowledge Bank" repository in Convai. The text files we developed from online sources were made available–one text file for each character. In informal testing we found that the default settings were sufficient to support reasonable responses that remained focused on the factual content of the artwork. Although Convai supports additional guardrails to prevent profanity and problematic requests, we chose not to use them to understand what visitors might discuss without censoring responses. We did not observe any abusive or profane conversations during the study.

In line with guidelines for museum labels (Australian Museum; Serrell, 2015), we designed the guide to provide concise, relevant information about the artworks. Voice settings and prompt structure was the same for all paintings, with the latter defined as:

> You are a museum curator explaining the painting '[Name of the painting]' by '[Name of the artist]'. You never discuss anything other than that painting or topics related to the painting. You only speak in short sentences between 20 and 30 words.

Which artwork the user is interacting with, and correspondingly which knowledge base is active, is determined by their position and orientation in space. When looking at a painting, users can press the trigger button of a controller to initiate a question. While the button is pressed, the users' speech is recorded and upon button release is sent to Convai. We provide audio feedback when the button is pushed and when it is released, allowing users to remain focused on the artwork. A spoken response is then generated by Convai and played back for the user. The response time of the system varied from 3 to 4 s.

3.2. Field study

We ran a field study in the lobby of the Museum of Ancient Art[4] in Aarhus, Denmark (see Fig. 2). This is a smaller museum, located on a university campus, that sees about 6–7 visitors per hour [5]and primarily exhibits artifacts from the antiquities, e.g. Greek and Roman sculptures, Egyptian mummies, etc. At the time of our study, a temporary exhibition was featured consisting of paintings inspired by ancient healing sanctuaries.

3.2.1. Participants

We recruited participants from among visitors as they were entering or leaving the museum. Visitors were eligible to participate if they were proficient in English and above the age of 18. Over the course of 10 days 22 people (9 males, 13 females, aged 20—68 years) agreed to participate. Participants had generally little experience with using HMDs (11× none, 7× a few times, and 4× occasion-ally or frequently) and with AI (10× minimal, 7× occasional, 5× frequently or regularly). Many of our participants visit a museum at least once or twice a year (6×), a few more times a year (8×), or even weekly or monthly (4×). However, we also had some participants who only rarely (3×) or very rarely (1×) visited museums. They mostly preferred written guides (17×), but also liked audio (14×) and personal (12×) guides. Only one participant reported preferring not to use any type of guide. Their motivations for visiting museums also varied, with all noting *"to discover or learn something new"* a many also checking *"to pay respect to history, culture, or heritage"* (18×), and *"to deepen knowledge related to work or interests"* (15×). Less common motivations included *"To* support *the experience of family or friends"* (10×), *"to find a peaceful, reflective space"* (7×), *"to visit an important or popular location"* (7×), and *"To connect with heritage or identity"* (6×).

3.2.2. Procedure

Participants first gave informed consent and then completed a pre-study questionnaire on demographics, museum visiting habits, and familiarity with HMDs and AI. We then introduced participants to the prototype and how to use it. They could then freely interact with the museum guide and we encouraged them to explore the prototype naturally for however much time they felt like.

Once they were done exploring the prototype, participants took off the head-set and we asked them to complete a post-study questionnaire (see Table 1). Afterwards, we conducted a short semi-structured interview (see Table 2). We also recorded how much time participants spend using the prototype, how much they engaged with the museum guide and for which artworks. In the application logs we also saved all queries and guide responses. We also took field notes as participants were using the prototype (see Table 3).

[1] https://assetstore.unity.com/packages/tools/behavior-ai/npc-ai-engine-dialog-actions-voice-and-lipsync-convai-235621.

[2] https://www.wassily-kandinsky.org/Yellow-Red-Blue.jsp.

[3] https://en.wikipedia.org/wiki/Wassily_Kandinsky, https://en.wikipedia.org/wiki/Nighthawks_(Hopper), https://en.wikipedia.org/wiki/The_Starry_Night, and https://en.wikipedia.org/wiki/A_Sunday_Afternoon_on_the_Island_of_La_Grande_Jatte.

[4] https://antikmuseet.au.dk/.

[5] https://antikmuseet.au.dk/museet/besoegsstatistikker.

Fig. 1. We used four virtual artworks in our prototype, all placed on virtual easels. Clock-wise from the top left these were: A Sunday Afternoon on the Island of La Grande Jatte by Georges Seurat (1886), Nighthawks by Edward Hopper (1942), The Starry Night by Vincent van Gogh (1889), and Yellow-Red-Blue by Wassily Kandinsky (1925).

Fig. 2. We ran the study in a 3 × 3m space in the corner of a museum lobby as indicated in the floorplan sketch.

3.2.3. Analysis

In line with the exploratory nature of the study, we analyzed the quantitative data with descriptive statistics. For the interview data, we used inductive thematic coding informed by grounded theory techniques (Strauss and Corbin, 1998). After transcription, the relevant excerpts were assigned short descriptive codes and compared using affinity diagramming. For example, statements about being "not entirely sure what to ask" were initially coded as *difficulty formulating questions*, while concerns about potentially "missing a lot" were coded as *risk of missing information*. These codes developed into the subthemes *Barriers to Spontaneous Questioning* and *Risk of Missing Important Information*, which were grouped under the higher-level theme *Active Information Seeking*. Codes and groupings were iteratively reviewed, with overlapping categories merged and distinct categories separated or renamed.

We used a deductive coding approach for the log data to identify the types of questions participants had for the museum guide. Adopting Barth et al.'s framework (Barth et al., 2020), we coded the intent behind each question as either:

Fact: Questions about who the artist was, when the artwork was created, its size, or where it has been exhibited.

Author: Questions about the artist's life, which art movement they were part of or stylistic influences.

Visual: Questions relating to colors, shapes, or the artwork's composition.

Style: Questions regarding the artistic style or movement to which the artwork belongs.

Context: Questions about the historical, social, or cultural context in

Table 1
Questionnaire answered by participants after the field study. All items were adapted from the MMGS (Othman et al., 2011), except for two additional questions we added (prefaced with *). All items rated on 5-point Likert scales; (R) = reverse-coded.

Scale dimension	Question
General usability	I would use the virtual guide again when I visit an exhibition
	The virtual guide was not a distraction
	The information given by the virtual guide was too lengthy (R)
	The information given by the virtual guide was too short (R)
	Using the virtual guide enhanced my exhibition visit The multimedia guide was not too complicated to use
Learnability and control	Learning to operate the virtual guide was easy
	I felt I was in control of the virtual guide
	The virtual guide presented information in an understandable manner
	I found it easy to understand the speech of the virtual guide
Quality of interaction	* The answers to my questions were relevant and were answered satisfactorily
	*I trusted that the information I was given was correct It was clear to me when the virtual guide was offering information and when I needed to ask for it
	I became unaware that I was even using any controls on the virtual guide.

Table 2
Questions used in semi-structured interview with participants after the field study.

Were your questions answered satisfactorily?
How does your experience with the virtual guide compare to traditional meth-ods of receiving information in museums, such as text labels or audio guides? How do you feel that the virtual guide affected your understanding of the artworks? In what ways?
What improvements would you suggest?
Anything else to add?

Table 3
Number of questions asked by the participants according to the intent types frame-work by Barth et al. (2020).

Fact	Artist	Visual	Style	Context	Meaning	Play
83	28	18	17	22	27	2

which the artwork was created.

Meaning: Questions related to intentions, meanings, or whys, and the stories possibly behind the people and elements depicted in the artwork.

Play: Utterances expressing humor, jokes, or playful interactions with the system.

Outside: Questions related to the conversational guide itself, its technology, or unrecognized utterances.

4. Results

All participants were able to engage in conversations with the guide to ask about the paintings. We now present usage statistics, summary of the questionnaire responses, overview of the questions posed by participants, and a thematic analyses of the interviews.

4.1. Usage and ratings

Participants spent on average 4.0 min (SD = 1.7) with the conversational museum guide. They interacted with 2.9 of the 4 paintings on average (SD = 0.9) and asked 9.2 questions on average (SD = 3.5) corresponding to approximately one question every 26 s.

The results of the multimedia guide scale (see Fig. 3) show overall positive average scores for general usability (M = 3.2, SD = 0.5), learnability and control (M = 4.1, SD = 0.5), and quality of interaction (M = 3.8, SD = 0.5). Particularly high here were ratings for the system being "not too complicated to use" (M = 4.4, SD = 0.6), that "learning to operate the guide was easy" (M = 4.5, SD = 0.7) and that the information was understandable (M = 4.5, SD = 0.51), indicating that the single-button design and natural language interface was effective at minimizing barriers to adoption. The speech output from the guide was also easy to understand (M = 3.7, SD = 1.2). The controls for the interaction, however remained present for the participants (M = 3, SD = 1.0) suggesting that the HMD and handheld may have influenced their interaction. Participants rated the relevance of the responses from the guide and their satisfaction with them positive (M = 4, SD = 0.6), as they did the overall trustworthiness of the information (M = 4.2, SD = 0.9).

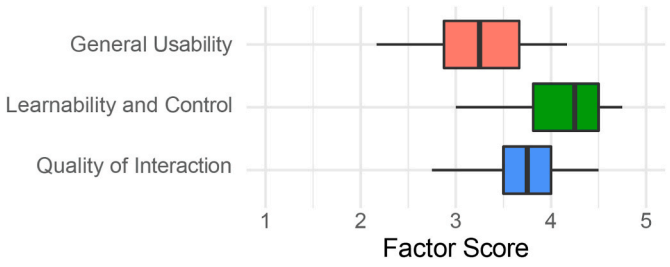

Fig. 3. Scores for each of the three factors of the questionnaire were overall positive.

4.2. Questions posed by participants

In order to understand what types of questions were asked by the participants to the guide, we analyzed the chat logs of each session: 197 questions users addressed to the guide. Each question was categorized according to the intent, folloring Barth et al. (2020). In this study, fact-based questions accounted for the overall majority of participant utterances, with 83 such questions directed at the museum guide (42%), indicating that users were primarily seeking factual information about the artwork. Questions also related to the artist (28), meaning (27) and context (22) were also popular, indicating interest in learning more about the artist, the deeper meaning behind, and context of the artwork. Questions related to visual elements (18), style (17), and play (2) were less common but still somewhat relevant for some participants. No questions were categorized as focusing on outside connections beyond the artist or the art (0).

4.3. Emergent themes from interviews

Our analysis of the interview transcripts uncovered several themes, with the six most prominent themes and various subthemes described in the following.

4.3.1. Theme 1: personalized interaction

The most common theme was the ability for participants to tailor the interaction to focus on the information they wanted to get from the guide, which reflects various examples of prior research (Vasic et al., 2024; Wakkary et al., 2008; Stock et al., 2007; Ardissono et al., 2012). Two distinct variations of this theme were identified and split in subthemes listed below.

Subtheme 1.1: Information Tailored to Interest. The ability to specify exactly what information to focus on about an artwork enables the user to access content relevant to their interests rather than being limited to the information chosen to display by the museum. This sentiment was expressed in statements such as:

P19: *"I think the smart thing about this is that you can ask about more specific things that interest you."*

P10: *"It can give me a bit more specific understanding of the different things. For example, I asked it what specifically the blue color represents and what about the white color in this picture? But if I was just reading about the picture, I imagine it would not mention what each color means."*

Another participant noted how the concept was particularly useful for visitors with specialized interests, without it being at the expense of those without such interests or preexisting knowledge about the subject:

P1: *"If it, for example, could answer those questions about Fauvism or whatever it was, then that would be super cool, because it would mean that I, as someone who maybe has a bit more of a special interest, could actually engage without it being at the expense of those who don't have a special interest. That's often the case in museums, where you come in, and everything has to cater to the most general audience possible. So if you come with a special interest, you can sometimes feel a bit talked down to, or that what's presented isn't particularly relevant."*

Subtheme 1.2: Selective Engagement with Content. Several participants expressed satisfaction with the ability to filter out the amount of unwanted information they had to consume in order to get the information

that they actually wanted. This allows the user to focus on the content which they are most interested in and avoid excessive or irrelevant content. This subtheme was represented in statements such as:

P18: *"Maybe you just want to know one thing, but then you have to go through a whole bunch of little texts with information you already know."*

P4: *"You could develop a bit more of a personal connection to it (the artwork), rather than just looking at it as a painting of such-and-such type from a certain year, created by so-and-so, where it's very basic information. Here, you could gain a bit more understanding, as well as perhaps the thought process or inspiration behind it, without necessarily having to read an entire book about Vincent van Gogh or whoever it was."*

One participant compared the guide to traditional audio guides, pointing out how they often require the user to take in unwanted information.

P11: *"I would say it's better than an audio guide because you can select the information you want."*

4.3.2. Theme 2: Active Information Seeking

As a counterpoint to the advantages of personalization, some participants expressed frustration with the fact that they needed to come up with questions unprompted on the spot. Traditional informative materials provide pre-curated comprehensive information, whereas the guide's question-driven nature requires users to actively request information. Challenges of the varied motivation levels (Sung et al., 2010) has been discussed in prior work with some participants desiring a more passive role (Schaffer et al., 2023), perhaps expecting the guide to manage the experience (Rubino et al., 2013). The active role as an information seeker rather than passive recipient presented a challenge for some users.

Subtheme 2.1: Barriers to Spontaneous Questioning. Some participants struggled to come up with questions for the guide spontaneously. Several participants required time to come up with a question and experienced moments of feeling "blank" or mentally blocked, which made it harder for participants to engage with the guide effectively.

During the interviews this concern was visible in statements such as:

P8: *"Here, the question is whether you ask the right questions so that you get a full understanding of the painter, or if you end up missing a lot because you're not entirely sure what to ask about. If you already have some knowledge, you can ask more in-depth questions. But if you don't, you might just stand there thinking, 'What should I ask now?"'*

P17: *"But I also think it's about knowing what kind of questions to ask because it can also be difficult to come up with questions."*

One participant pointed out how the appropriateness of the guide might also depend on the visitor's state of mind at the time.

P12: *"I think it really depends on my mood that day. How much energy I feel I have for those kinds of thoughts. For example, today I have a headache, so I might not be at my best today."*

Subtheme 2.2: Risk of Missing Important Information. Some participants expressed concern that they might miss out on important information if they rely only on user-initiated questions. The risk of the user realizing what is important to ask about might result in information being overlooked, leaving gaps in the information provided. Traditional informative materials proactively deliver well structured, pre-curated content which ensures that the user is presented with a base level of information regardless of their input. This perspective was expressed in quotes such as:

P8: *"The question is whether you're asking the right questions to cover the whole artist, or if you miss a lot because you're not entirely sure what you should be asking about."*

P15: *"I get a better understanding of things when I read about them. But with that one, I only get what I ask about. So maybe it's something about not necessarily knowing what to ask about?"*

P18: *"But I can just feel that maybe I could miss something—something I forget to ask about, which might have been written on a sign, for example."*

4.3.3. Theme 3: public self-consciousness

Some participants expressed concerns about feeling self-conscious about having to express themselves vocally in the public space to interact with the guide, especially in the presence of other visitors. Participants expressed discomfort with the idea of asking the guide questions out loud due to the fear of being judged by others. This is a similar concern as identified in previous research showing that some museum visitors feel uncomfortable asking tour guides questions which might be perceived as stupid questions (Thiel and Bernhardt, 2023). This concern was echoed in statements made such as:

P6: *"Well, it's a bit more boundary-pushing in a way, but it's also like … Yes, I mean, it's that you have to walk and talk while being very aware that there are other people around, in that sort of way."*

P18: *"You're usually quiet in a museum; you don't really talk much—just whisper to each other. But now, all of a sudden, I have to stand and talk to this fictional painting in a room where people can walk by, and there are staff sitting at the front desk and so on."*

In contrast, some participants felt more free to ask questions of the guide than they would have if there was a human guide.

P2: *"Also, in relation to tour guides and group tours in a museum, I've been on a tour where I had some questions I wanted to ask. But I couldn't bring myself to say them because I didn't want to look stupid or ask a silly question. […] So it's like having your own personal guide."*

4.3.4. Theme 4: language accessibility

Communicating with the visitor in their preferred language has been an aspiration identified in various technology-enhanced guides (Trichopoulos, 2023; Thiel and Bernhardt, 2023). While the majority of participants interacted with the Virtual Agent fluently in English, six participants (4 above 50 years old) expressed a desire for the guide to support other languages, in particular Danish. This perspective was expressed in statements such as:

P7: *"If it were going to work it would need to be in more different languages.*

At least Danish and English here."

P21: *"I would have been able to use this much better if it had been in Danish.*

That's because I'm so old."

P11: *"So, I was thinking, like, which time period it was from, artistically speaking. But I could not really formulate that in English. So I just needed a bit of time to think about it."*

4.3.5. Theme 5: engagement

Several participants report feeling more engaged when using the prototype compared to traditional museum materials. This phenomenon aligns with the broader goal of museums incorporating interactive technologies to increase visitor engagement (Pallud, 2017; Hammady et al., 2020). The theme was separated into two different subthemes.

Subtheme 5.1: Technology Engagement. Comparing the guide to more traditional museum materials, some participants described feeling more engaged with the content when using the prototype due to it feeling less heavy and more interactive.

P10: *"It's not quite as heavy in the same way as the traditional can be. So it's a slightly more fun way to do it."*

P12: *"This might be a bit more selective because you can interact more with things since you have to ask questions yourself and think about them. In a way, you also engage yourself in the process."*

One participant reported this as particularly effective for visitors in museums which might not appeal to their prior interests — in this case, for her, an art museum.

P13: *"When it's an art museum, I need to be drawn into it more because it's not something that naturally captivates me. So, in that case, it would definitely be a clear advantage."*

Subtheme 5.2: Novelty. It was noted by some participants that the novelty of the technology used for the prototype made their experience more appealing. The novelty of the technology was confirmed by

reviewing the demographic data related to participants prior experience using a HMD and AI.

P10: *"I don't know if I would chose it every time (the guide), but it is fun to try something new."*

P12: *"I think, I believe it was cool to try something new."*

4.3.6. Theme 6: decreased immersion

Prior research suggests that guides might create a barrier to immersion in the museum experience (Sung et al., 2010; Lee, 2017; Clarke et al., 2021). Despite appreciating the assistance provided by the guide, some participants found that it created a barrier between them and their sense of immersion within the museum environment. These participants felt that it disrupted the sensory and atmospheric aspects of the museum visit. P7: *"But I also think, in a way, that it can be a bit stressful to have that technical, technological aspect between yourself and the artwork, instead of just* walking around and experiencing the artwork directly."

P9: *"For me, it disrupts the experience. I should be allowed to use my senses."*

5. Discussion

Our field study has shown how museum visitors have a range of needs and thus also interacted with the guide in different ways. This necessitates personalization of museum experiences, which aligns with findings from previous work. We have also found that there is a need for further refining how users can interact with such guides and the barriers they currently face. Furthermore, how best to test these guides connects to this and improvements in this regard also are needed. In the following, we discuss these aspects related to our conversational museum guides in more detail.

5.1. Balancing personalization and curation

The majority of participants appreciated the ability to customize the information they were provided, which is consistent with prior research (Vasic et al., 2024; Wakkary et al., 2008; Stock et al., 2007; Ardissono et al., 2012). The findings of this study suggest the potential of LLM-powered guides to transform how museum visitors access information, shifting from traditional static materials to more dynamic and interactive methods, which can adjust to the unique needs of individual visitors. A promising direction in personalization, "recipient design" involves tailoring the CA to the user's knowledge, prior difficulty in understanding, and prior exposure to a reference (An et al., 2021). This technique requires logging the conversation and adapting to the explicit questions and tracking the progress throughout the interaction. While our system did log the conversation, an important next step would be to build and test the effectiveness of these techniques in adapting to the user. In previous research involving an app-based CA museum guide (Barth et al., 2020), questions in their studies were more focused on the *meaning* of the artwork, yet in our study, there were far more *fact* related questions. In the app-based work, when participants approached a painting, they were presented with an onscreen label showing a thumbnail of the artwork, the title, artist, year, location, and gallery location, which already provides essential factual details to the visitor and this likely prompted them to ask deeper questions about the *meaning* of the artwork (Barth et al., 2020). The design choice of presenting the factual details may give the user a frame of reference from which to ask deeper questions, however, for some visitors, the date and title may not be the most important details needed to appreciate the work. In our system, no facts were front-loaded and users had to explicitly query for them.

Our study uncovered the tension and contradictory demands from museum visitors that deserves further investigation. Museum visitors may want technology to deliver a personalized experience, yet at the same time, expect some level of curation for ease of interaction and to ensure they do not miss out on key details. The potential of the latest generative AI models includes retrieval of information from external sources, reasoning, and agentic processes that could ensure information is up to date (Trichopoulos et al., 2023). However, the role of the human curator involves taste and style, which is beyond the reach of AI for now. Perhaps a fruitful direction is to embrace the curators and involve them more deeply in the process of designing guides and to place them in the driver's seat to not just filter and fine-tune (Vasic et al., 2024), but to develop their own guides and to experiment with how to capture the interest of the visitors and stage meaningful learning experiences. This is not a straightforward endeavor and recent work has explored how to unveil information through interaction (Sadiri et al., 2024), conveying factual information is not always simply a task of selecting and regurgitating facts. Especially in a museum setting, human guides often weave in a story about the painting or artist and unfold the facts. Our guide serves as a proof-of-concept, supporting the relevance and value of conducting further research in this area.

5.2. Barriers to interaction

Participants encountered several challenges when interacting with the guide. Some found it difficult to generate creative or reflective questions, often defaulting to fact-based inquiries. In our study, there was no time pressure intended toward the participants, yet they may have felt obligated to ask clever questions and fill the time. Coming up with questions does involve mental effort and active participation. This indicates the potential for supportive strategies to get the conversation going, such as framing interactions as a mystery game, to enhance engagement (Simarro et al., 2005).

Uncertainty about what to ask also raised concerns from the participants about their fear of missing key information, a phenomenon noted in prior research on the anxiety of self-directed learning in public spaces (Pallud, 2017). Additionally, some participants felt uneasy interacting vocally in public, with a few suggesting a text-based option to allow for more discreet communication, to avoid being the center of attention or asking silly questions (Thiel and Bernhardt, 2023). Other recommendations included displaying supplementary text (keywords, prompts, or full descriptions) alongside artworks to spark interest and aid question formulation, as well as offering multilingual support to accommodate diverse visitor needs.

5.3. Evaluation methods for guides

Our study approach involved the use of logged data, the modified MMGS questionnaire, and contextual interviews, which provided valuable insights into the usability, learnability, and interaction quality of the guide. The original MMGS is a highly specialized tool for measuring the effectiveness of multimedia guides designed for a museum context (Othman et al., 2011), yet does not account for measuring the experiences with advanced technologies such as AI. Evaluation of museum experiences is a persistent challenge with recommendations to conduct mixed-methods approaches (Hornecker and Ciolfi, 2019), ethnography, or observations, which might yield deep insights and detailed accounts of the experience, yet are often expensive, time consuming and difficult to conduct (Roussou and Katifori, 2018).

Considering that interactive experiences are supported by digital technologies that can record behavior and sample user experience while embedded in the context of use, various probes (Graham et al., 2007) and sampling methods have gained in popularity. With advances in generative AI, there are beginning to be examples of museum guides that can handle changes in context such as location and time (Vasic et al., 2024) and adapt the behavior of the guide. With increasingly sophisticated models and techniques such as agentic AI and chain of thought reasoning (Yao et al., 2023), there is potential to build systems that learn from the conversational interactions and adapt. Similar to technology probes (Graham et al., 2007), the agent log can provide insights about the user experience out in the wild and away from the constraints of a

controlled lab study. How the virtual agents may adapt requires careful consideration and additional work is needed including more involvement from curators.

5.4. Design implications for conversational museum guides

Taken together, the interaction logs, questionnaire results, and interviews suggest several implications that can inform future guidelines for conversational museum guides.

Support *question formulation*. Although participants found the guide easy to use, some were uncertain about what to ask, and most questions were factual. Guides could therefore offer optional example questions, suggested topics, or context-sensitive prompts. These should support interaction without limiting visitors' freedom to pursue their own interests.

Balance personalization and curation. Participants valued receiving information tailored to their interests, but were concerned that they might miss important content. Future systems could combine visitor-led questioning with a brief curator-selected introduction or occasional suggestions about unexplored topics.

Provide flexible interaction modes. Speaking aloud in a public museum was uncomfortable for some participants, while language also affected question for-mulation. Text input, selectable prompts, subtitles, discreet audio, and multilingual support could make interaction more accessible and appropriate for different visitors and settings.

5.5. Future directions and considerations

Participants rated the guide highly in terms of trust, an encouraging result for museums considering AI-powered guides. This positive reception underscores the developers' responsibility to ensure the accuracy of the information provided, as even modern AI systems risk propagating outdated or incorrect data (Currie, 2023).

In addition, our participant demographic indicated that most participants had little experience with AI technologies like ChatGPT, with some noting the novelty as a fun aspect of the study. This suggests that perceptions of guides may shift as advanced AI becomes more common. Ethical considerations remain equally important, as the drive toward increased accessibility and interaction is met with resistance from those who might fear that AI-guided systems could compromise institutional integrity and devalue curatorial expertise (McCall and Gray, 2014). Future research might explore combining advanced technology with human expertise, supporting rather than replacing human guides (Antoniou et al., 2021).

Furthermore, many museum frameworks emphasize the importance of social interaction, current designs focusing on individual device interactions often fail to support group dynamics, a shortcoming that future research should address to better accommodate groups such as families and school field trips (Grinter et al., 2002).

6. Conclusion

This study examined the integration of a modern conversational agent as a virtual museum guide to enhance visitor engagement through natural langauge interactions about paintings. Leveraging an LLM supported conversational agent platform, the prototype enabled personalized, interactive experiences that overcome the limitations of static, one-size-fits-all informational materials and labels. Evaluations with real museum visitors—each with unique interests—revealed positive attitudes regarding usability, learnability, and inter-action quality. Qualitative insights underscored the benefits of personalization and selective engagement, while also identifying challenges such as formulating effective questions, addressing potential information gaps, and overcoming self-consciousness in public verbal interactions. Overall, this work contributes valuable insights into how AI-driven guides could transform museum experiences by emphasizing user perceptions alongside technical performance.

CRediT authorship contribution statement

Tobias Tue: Conceptualization, Data curation, Formal analysis, Investigation, Software, Writing – original draft, Writing – review & editing. **Henning Pohl:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Timothy Merritt:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

An, S., Moore, R., Liu, E.Y., Ren, G.-J., 2021. Recipient design for conversational agents: tailoring agent's utterance to user's knowledge. In: Pro-Ceedings of the 3rd Conference on Conversational User Interfaces, CUI '21. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3469595.3469625.

Antoniou, A., Vayanou, M., Katifori, A., Chrysanthi, A., Cheilitsi, F., Ioan-nidis, Y., 2021. "real change comes from within!": towards a symbiosis of human and digital guides in the museum. J. Comput. Cult. Herit. 15 (1). https://doi.org/10.1145/3465557.

Ardissono, L., Kuflik, T., Petrelli, D., 2012. Personalization in cultural heritage: the road travelled and the one ahead. User Model. User-Adapted Interact. 22 (1–2), 73–99. https://doi.org/10.1007/s11257-011-9104-x.

Australian Museum, Writing text and labels, https://australian.museum/learn/teachers/learning/writing-text-and-labels/, ac-cessed: 2023-December-19 (n.d.).

Banerjee, A., Satish, A., Wörndl, W., 2024. Enhancing tourism recommender systems for sustainable city trips using retrieval-augmented generation. arXiv:2409.18003. URL. https://arxiv.org/abs/2409.18003.

Barth, F., Candello, H., Cavalin, P., Pinhanez, C., 2020. Intentions, meanings, and whys: designing content for voice-based conversational museum guides. In: Proceedings of the 2nd Conference on Conversational User Interfaces, CUI '20. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3405755.3406128.

Bin, Y., Shi, W., Ding, Y., Hu, Z., Wang, Z., Yang, Y., Ng, S.-K., Shen, H.T., 2024. Gallerygpt: analyzing paintings with large multimodal models. In: Proceedings of the 32nd ACM International Conference on Multimedia, MM '24. Association for Computing Machinery, New York, NY, USA, pp. 7734–7743. https://doi.org/10.1145/3664647.3681656.

Booth, B., 1998. Understanding the information needs of visitors to museums. Mus. Manag. Curatorship 17 (2), 139–157. https://doi.org/10.1080/09647779800301702.

Chen, B., Wen, R., Tan, S., Li, Y., 2025. Exploring user preferences for museum guides: the role of chatbots in shaping interactive experiences. In: Proceed-Ings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '25. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3706599.3720067.

Clarke, L., Hornecker, E., Ruthven, I., 2021. Fighting fires and powering steam Locomotives: distribution of control and its role in social interaction at tangible interactive museum exhibits. In: Proceedings of the 2021 CHI confer-ence on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445534.

Currie, G.M., 2023. Academic integrity and artificial intelligence: is chatgpt hype, hero or heresy?. In: Seminars in Nuclear Medicine, 53. Elsevier, pp. 719–730 aI trust reliability.

Duguleană, M., Briciu, V.-A., Duduman, I.-A., Machidon, O.M., 2020. A virtual assistant for natural interactions in museums. Sustainability 12 (17). https://doi.org/10.3390/su12176958.

Falk, J.H., Dierking, L.D., 2018. Learning from Museums, second ed. Edition. Rowman & Littlefield, Lanham.

Gaia, G., Boiano, S., Borda, A., 2019. Engaging museum visitors with AI: the case of chatbots. In: Unknown Book Title. Springer International Pub-lishing, Cham, pp. 309–329. https://doi.org/10.1007/978-3-319-97457-6_15.

Graham, C., Rouncefield, M., Gibbs, M., Vetere, F., Cheverst, K., 2007. How probes work. In: Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces, pp. 29–37.

Grinter, R.E., Aoki, P.M., Szymanski, M.H., Thornton, J.D., Woodruff, A., Hurst, A., 2002. Revisiting the visit: understanding how technology can shape the museum visit. In: Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work, pp. 146–155 social isolation.

Hammady, R., Ma, M., Strathern, C., Mohamad, M., 2020. Design and development of a spatial mixed reality touring guide to the Egyptian museum. Multimed. Tool. Appl. 79 (5), 3465–3494. https://doi.org/10.1007/s11042-019-08026-w.

Hornecker, E., Ciolfi, L., 2019. Human-Computer Interactions in muse-ums. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-031-02225-8.
Lee, S.J., 2017. A review of audio guides in the era of smart tourism. Inf. Syst. Front. 19, 705–715. https://doi.org/10.1007/s10796-016-9666-6.
Lee, U., Jeon, M., Lee, Y., Byun, G., Son, Y., Shin, J., Ko, H., Kim, H., 2024. Llava-docent: instruction tuning with multimodal large language model to support art appreciation education. Comput. Educ.: Artif. Intell. 7, 100297. https://doi.org/10.1016/j.caeai.2024.100297. https://www.sciencedirect.com/science/article/pii/S2666920X24001000.
Lin, W., Hu, X., 2026. Museforge: enhancing creative learning in digital museum education with generative ai. In: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, CHI '26. Association for Com-puting Machinery, New York, NY, USA. https://doi.org/10.1145/3772318.3791173.
Maniscalco, U., Minutolo, A., Storniolo, P., Esposito, M., 2024. Towards a more anthropomorphic interaction with robots in museum settings: an experimental study. Robot. Autonom. Syst. 171, 104561. https://doi.org/10.1016/j.robot.2023.104561.
McCall, V., Gray, C., 2014. Museums and the 'new museology': theory, practice and organisational change. Mus. Manag. Curatorship 29 (1), 19–35 old Museology.
Othman, M.K., Petrie, H., Power, C., 2011. Engaging visitors in museums with technology: scales for the measurement of visitor and multimedia guide ex-perience. In: Human-Computer Interaction–INTERACT 2011: 13Th IFIP TC 13 International Conference, Lisbon, Portugal, September 5-9, 2011, Proceedings, Part IV 13. Springer, pp. 92–99.
Pallud, J., 2017. Impact of interactive technologies on stimulating learning expe-riences in a museum. Inf. Manag. 54 (4), 465–478. https://doi.org/10.1016/j.im.2016.10.004.
Roussou, M., Katifori, A., 2018. Flow, staging, wayfinding, personalization: eval-uating user experience with Mobile museum narratives. Multimodal Technol. Interact. 2 (2), 32. https://doi.org/10.3390/mti2020032.
Rubino, I., Xhembulla, J., Martina, A., Bottino, A., Malnati, G., 2013. MusA: using indoor positioning and navigation to enhance cultural experiences in a museum. Sensors 13 (12), 17445–17471. https://doi.org/10.3390/s131217445.
Sadiri Javadi, V., Trippas, J.R., Flek, L., 2024. Unveiling information through narrative in conversational information seeking. In: Proceedings of the 6th ACM Conference on Conversational User Interfaces, CUI '24. As-sociation for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3640794.3665884.
Schaffer, S., Ruß, A., Gustke, O., 2023. User experience of a conversational user interface in a Museum. In: Artsit, Interactivity and Game Creation. Springer, Cham, pp. 215–223. https://doi.org/10.1007/978-3-031-28993-4_16 iSSN: 1867-822X. https://link.springer.com/chapter/10.1007/978-3-031-28993-4_16.
Schaffer, S., Schwaetzer, E., Ruß, A., Gustke, O., 2024. Chatbot in the Museum–A field study of user experience and modality usage. In: Elektronische Sprachsignalverarbeitung 2024, Tagungsband Der 35. Konferenz, Regens-burg, 6.-8. März 2024. TUDpress, pp. 14–21. https://doi.org/10.35096/othr/pub-7075.
Serrell, B., 2015. Exhibit Labels: an Interpretive Approach, second ed. Row-man & Littlefield, Lanham, MD. Ch. 3, label Lengths.
Simarro Cabrera, J., Muñoz Frutos, H., Stoica, A.G., Avouris, N., Dim-itriadis, Y., Fiotakis, G., Liveri, K.D., 2005. Mystery in the museum: collaborative learning activities using handheld devices. In: Proceedings of the 7th International Conference on Human Computer Interaction with Mo-bile Devices & Services. ACM, Salzburg, Austria, pp. 315–318. https://doi.org/10.1145/1085777.1085843.
Stock, O., Zancanaro, M., Busetta, P., Callaway, C., Krüger, A., Kruppa, M., Kuflik, T., Not, E., Rocchi, C., 2007. Adaptive, intelligent presentation of information for the museum visitor in PEACH. User Model. User-Adapted Interact. 17 (3), 257–304. https://doi.org/10.1007/s11257-007-9029-6.
Strauss, A., Corbin, J., 1998. Basics of Qualitative Research Techniques.
Sung, Y.-T., Chang, K.-E., Hou, H.-T., Chen, P.-F., 2010. Designing an electronic guidebook for learning engagement in a museum of history. Comput. Hum. Behav. 26 (1), 74–83. https://doi.org/10.1016/j.chb.2009.08.004.
Thiel, S., Bernhardt, J.C. (Eds.), 2023. AI in Museums: Reflections, Perspectives and Applications, first ed. transcript Ver-lag, Bielefeld, Germany. https://doi.org/10.14361/9783839467107. Vol. 74 of Edition Museum. https://www.transcript-open.de/isbn/6710.
Tokuno, K., Kusunoki, F., Inagaki, S., Mizoguchi, H., 2024. Talkative museum: augmented reality interactive museum guide system towards collaborative child-parent-specimen interaction. In: Proceedings of the 23rd An-nual ACM Interaction Design and Children Conference, IDC '24. Associ-ation for Computing Machinery, New York, NY, USA, pp. 754–758. https://doi.org/10.1145/3628516.3659389.
Trichopoulos, G., 2023. Large language models for cultural heritage. In: Pro-Ceedings of the 2nd International Conference of the ACM Greek SIGCHI Chapter, CHIGREECE '23. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3609987.3610018.
Trichopoulos, G., Konstantakis, M., Caridakis, G., Katifori, A., Kouk-ouli, M., 2023. Crafting a Museum guide using ChatGPT4. Big Data and Cognitive Computing 7 (3). https://doi.org/10.3390/bdcc7030148.
Vasic, I., Fill, H.-G., Quattrini, R., Pierdicca, R., 2024. LLM-aided Museum guide: personalized tours based on user preferences. In: De Paolis, L.T., Arpaia, P., Sacco, M. (Eds.), Extended Reality. Springer Nature, Switzerland, Cham, pp. 249–262.
Wakkary, R., Muise, K., Tanenbaum, K., Hatala, M., Kornfeld, L., 2008. Situ-ating approaches to interactive Museum guides. Mus. Manag. Curatorship 23 (4), 367–383. https://doi.org/10.1080/09647770802517423 publisher: Routledge.
Wang, T., Deterding, S., 2026. Conversational agents as relational mediations for glam: a conceptual framework. In: Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems, CHI EA '26. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3772363.3798504.
Wei, Q., Yang, M., Wang, J., Mao, W., Xu, J., Ning, H., 2024. Tourllm: enhancing llms with tourism knowledge. arXiv:2407.12791. URL. https://arxiv.org/abs/2407.12791.
Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., Cao, Y., 2023. React: synergizing reasoning and acting in language models. In: International Conference on Learning Representations (ICLR).