Human-Machine Trust in "Cyber Guides": A User Study of Large Language Model-Based Intelligent Guides in the West Lake Museum Context

Jiayu Yao

ZheJiang University of Technology, Hangzhou, China
3100941568@qq.com

Abstract. As generative AI agents increasingly serve as cultural intermediaries in museums, empirical research on tourists' trust in these "cyber guides" remains scarce. Drawing on human-computer trust and parasocial interaction theories, this mixed-methods study (surveys and interviews) at the West Lake Museum investigates the dimensions, determinants, and behavioral outcomes of tourist trust in AI guides. Findings reveal that such trust is inherently selective. Alongside traditional dimensions——competence, integrity, and benevolence——affective trust is emerging. Furthermore, human-AI trust is fragile; a single negative interaction causes severe, costly-to-repair trust violations. This paper refines and expands the application of human-machine trust theory in cultural tourism contexts, offering concrete recommendations for optimizing AI-guided tour designs in museum institutions. It serves as a reference for enhancing the efficiency of cultural management and cultural services in the era of AI.

Keywords: AI Agent, Human-Machine Trust, Intelligent Guide, Para-social Interaction

1. Introduction

The AI agents that are run by large language models are becoming more than mere chat bots but rather intelligent beings with proactivity, interactivity, and flexibility. Museums have become a major area of application of AI agents due to their role as the most popular providers of cultural services to the populace [1]. Starting in 2024, the Palace Museum has started a mini-program entitled "AI Tour Guide", which offers real-time guidance and answers to visitors depending on their location: The next-generation Shanghai Museum has deployed intelligent navigational screens that can perform multi-round conversations and customised routes; In January 2026, Hangzhou West Lake Museum added a "AI Image Recognition" capability, enabling visitors to learn more about culture through taking pictures of cultural heritage. Statistics by the China Museum Association indicate that by the end of 2025, more than 40 of the highest-ranking museums across the country had deployed AI-assisted navigation applications, which is an indication that intelligent navigation moves beyond the pilot phase to actual use. These "cyber tour guides" represent the fact that museum navigation services are changing towards intelligence instead of digitalization.

Nevertheless, the way trust is built up, sustained, or in some cases, dismantled, when tourists engage with anthropomorphic AI tour guides becomes an urgent practical and theoretical problem that needs to be researched. This paper addresses AI-assisted smart tour guiding at the West Lake Museum and raises three research questions:

Q1: In what dimensions is tourists' trust in AI-powered tour guidance manifested?

Q2: Which elements make up this trust? What role does the smoothness of interactive experiences, the accuracy of the information and the degree of AI anthropomorphism play in the formation of trust?

Q3: How does trust impact tourists satisfaction and return visits once trust has been created?

2. Literature review and theoretical framework

The idea of human-machine trust was introduced by Muir for the first time in a systematic way, asserting that the essential aspect of human-machine trust is how people feel about machines being able to help them reach their aims in unpredictable conditions. The three dimensions of human-machine trust were later defined and formed by Mayer and others as competence, integrity, and goodwill, which have become the standard of analysis in the field [2].

With the advent of the generative AI era, AI systems have started to acquire some of the qualities of conversation, memorization, and emotional presentation, changing the nature of human-machine interactions to not just operate but communicate in real sense [3]. Trust was revealed as one of the main factors influencing the use of technology, especially in the case of service-related issues involving AI through a systematic analysis of 23 articles published between 2021 and 2024 [4]. Hou et al. interviewed 385 users of museum service robots and discovered that trust beliefs also serve as mediators in the relationship between the service quality, perceived risk, perceived professionalism, and the trust intention, which is very pertinent to the current research [5].

The AI agents are the new frontier of modern-day artificial intelligence growth,which was formulated by Wooldridge and Jennings, who pointed out four main features: autonomy, social skills, responsiveness, and proactivity [6]. The museum tour guide services have undergone three stages of development: manual guidance, digital guidance and intelligent guidance [7]. As stated by Derda et al. , museum AI studies should not be seen merely as technical but rather human-oriented AI with a view of the ethical aspect, including the issues of transparency, inclusiveness, and human-orientation. The suggested analytical framework is constructed on the basis of the above literature review.

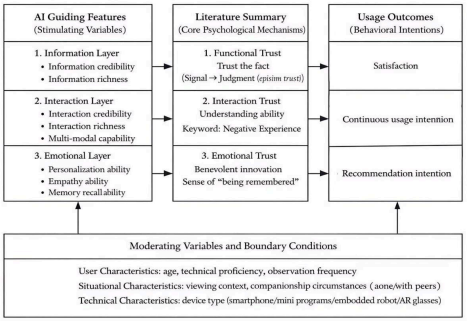

Figure 1. Research framework: the formation mechanism of human-machine trust in museum AI guides

In this framework AI-assisted navigation features are used as the dependent variable, which includes affective trust to determine if users create quasi-social emotional relationships with AI guide, and also explore the moderating role of negative experiences on trust.

3. Research design

3.1. Case selection

The case site of this study is the Hangzhou West Lake Museum. This decision was made because: the museum introduced the intelligent guide function of the so-called AI Photo Recognition in February of 2026, which is an example of an AI agent application that is closely related to the topic of research [8]; the museum is situated in the center of the West Lake scenic area and has received more than 23 million visitors since its opening 20 years ago with its visitors coming in different ages, areas and education levels [9];the position of the researcher in Hangzhou makes it possible to conduct several return flights in the form of field research, on-site observation and additional interviews.

3.2. Questionnaire design and distribution

Table 1. Questionnaire structure

Section	Content	Question	Instructions
1	Basic Information	5	Age, Educational Background, Visit Frequency, Technical Proficiency
2	Behavioral Usage	3	Duration, Purpose, Frequency
3	Trust Measurement	12	Adapted from Lee & See 、 Mayer et al.
4	Satisfaction and Usage Intention	4	Overall Satisfaction, Intention to Continue Use, Intention to Recommend

Table 2. Trust measure scale

Dimension	Item Number	Item Content
Functional Trust	FT1	The artifact information identified by AI is accurate.
	FT2	The AI answer content is trustworthy
	FT3	The AI-generated explanation was helpful to me.
Interactive Trust	IT1	AI can understand the questions I pose
	IT2	The AI conversation flows smoothly
	IT3	AI responds with remarkable speed
	IT4	Even if the AI occasionally makes mistakes, I am willing to continue using it.
Emotional Trust	ET1	During use, I felt respected
	ET2	This AI tour guide acts like a friend accompanying me on my tour.
	ET3	AI remembers my preferences, making me feel at home
	ET4	When I am done with the museum, I cannot resist nostalgia of this AI.
	ET5	I'd like to recommend this AI tour guide to my friends

3.3. Distribution method and sample retrieval

Researchers carried out random intercept surveys at the exit of Hangzhou West Lake Museum between March 1 and March 10,2026. The screening criteria were: (1) having completed most or all of the visit to the museum; (2) having used the AI-powered photo recognition technology; (3) being willing to take part in the survey. The questionnaire was administered on-site in hard copy form, which took between 5 and 8 minutes to complete. There were a total of 200 questionnaires handed out and 176 valid responses were gathered, hence, the effective response rate is 88%.

Table 3. Basic demographic distribution of the sample

Variable	Category	Frequency	Percentage
Age	18-30	102	58%
	31-45	49	27.8%
	Over 46	25	14.2%
Educational Background	High school and below	21	11.9%
	Associate or Undergraduate	125	71.0%
	Master or above	30	17.1%
Visit Frequency	Once or twice a year	111	63.1%
	More than 3 times per year	37	21.0%
	First visit	28	15.9%
Technical Proficiency	Frequently use	118	67.0%
	Occasionally use	48	27.3%
	Rarely use	10	5.7%

3.4. Interview implementation and findings

Once the questionnaire was done, 12 participants were chosen to be interviewed in semi-structured manner. Having received their permission, all the interviews were recorded and lasted around 235 minutes, which gave about 32,000 words transcript.

Table 4. The basic information of the respondents

Number	Age	Educational Background	Frequency of Utilization	Satisfaction	Interview Duration
P01	24	Bachelor	First time	High	18min
P02	32	Associate	Occasionally	Medium	22min
P03	28	High school	First time	High	15min
P04	45	Associate	First time	Low	20min
P05	21	Bachelor	First time	High	17min
P06	38	Ph.d	Repeated use	High	25min
P07	29	Bachelor	First time	Medium	16min
P08	52	High school	First time	Low	19min
P09	23	Bachelor	Occasionally	High	21min
P10	35	Master	First time	Medium	18min

Table 4. (continued)

P11	26	Bachelor	First time	High	23min
P12	41	Associate	Occasionally	Medium	20min

Table 5. Usage ratio

Usage	Percentage
Understand exhibit details	82.4%
Recognize unfamiliar objects	67.0%
Ask for directions/Find area	31.3%
Listen to the extended story	28.4%
Just curious to give it a try	18.2%

3.5. Dimensions of human-machine trust performance

Functional Trust

78.4% of users accept the correctness of AI recognition, however, in case of AI making subjective evaluations, the rate of full adoption of users reduces to 41.5%.

"This bowl is of the Song Dynasty, as it was stated to me. I believe that would be acceptable. However, when it has to do with choosing the most beautiful one, I still need to view it personally. "- Respondent 3, man, 24

User trust in AI is selective — factual outputs receive high trust, while judgmental outputs encounter low trust. Human-machine trust is not simply a matter of trusting or distrusting overall; rather, users adopt different trust strategies depending on the type of information.

Interactional Trust

The ability of AI to comprehend was recognized by 72.2% of the users, but 65.3% of them indicated that if there is one wrong answer it would decrease their level of trust.

"On my first attempt, I posed three queries, and all of them were correctly answered. I found it fairly credible. Subsequently, I inquired about the purpose of the bronze artifact and it responded: This is a fine piece of cultural heritage. I thought it was just trying to avoid me, and after that, I did not want to use it much more." - Respondent 7, female, 28

Human-machine trust is highly fragile——a single negative experience can instantly destroy trust accumulated through multiple positive experiences, and the cost of repair is extremely high. This distinguishes it from interpersonal trust, revealing the asymmetry of human-machine trust: building trust requires repeated positive feedback, while a single mistake can break it.

Emotional Trust

The agreement of 34.7 percent of users was that the AI was like a friend escorting them through the tour which increased to 42.3 percent in those aged 18-25.

"When it announced my name, I was somewhat shocked but later on, it seemed to be rather cozy. After that I inspected other exhibitions and I asked myself if it would recognize me once more." - Respondent 5, woman, 21

"Going to the museum by yourself can be rather boring most of the time. The feeling is much improved when you have someone to converse with." - Respondent 9, male, 23

Users can indeed develop an emotional connection with AI similar to interpersonal interactions ——the para-social relationship is extending from TV characters to AI. Therefore, human-machine

trust may not only include the classic three dimensions of ability, integrity, and benevolence, but may also have an emotional trust dimension. However, this connection remains relatively rudimentary and shows significant individual differences.

3.6. The influence of trust

Table 6. Correlation coefficient between trust dimension and satisfaction/ recommendation willingness

Trust Dimension	Correlation Coefficient with Satisfaction	Correlation Coefficient with Recommendation Willingness
Interactive Trust	r=0.62	r=0.51
Emotional Trust	r=0.48	r=0.45
Functional Trust	r=0.43	r=0.38

In case of museum AI guides, interactional trust plays a bigger role in user experience compared to functional trust. This contradicts the design intuition that values technical accuracy above anything else- users are ready to accept the presence of inaccurate information, but they cannot accept interaction lags or irrelevant answers. The most significant predictor of recommendation intention is emotional trust. Those users that form a para-social relationship with the AI are more likely to promote it among other people, which means that positive perception of warmth and humanization of the design may be translated into word-of-mouth marketing- the AI is not an instrument, but a kind of a friend.

4. Result

According to the research above, the human-machine trust in the museum setting has a selective nature. The trust in the accuracy of the outputs of the AI by users is very high whereas their trust in judgmental outputs is much lower. Users use various trust mechanisms due to the kind of information, i.e., even with the same AI, users allocate different levels of trust depending on the content type of its output. It shows that human-machine trust is very context-sensitive and content-sensitive.

Secondly, human-machine trust consists of three factors, which are functional, interactive, and emotional. Functional trust is also the basis, interactive trust is the protection, and emotional trust is a new feature. About 34.7% of users said that they had emotions similar to those during interpersonal interactions, e.g., feeling like a friend who was guiding them through a tour.

Thirdly, it takes only one negative experience to destroy the already existing trust. As many as 65.3 % of users stated that one wrong answer has a negative impact on trust, and the conducted interviews showed that even after the further exchanges get back to normal, users could still be suspicious about correct information. Thus, the fault tolerance in AI design is not a benefit but a necessity, designers need to create the trust recovery methods to restore user confidence with the help of future interactions in case of mistakes.

The fourth is trust has a positive correlation in relation to satisfaction and willingness to recommend. Interaction-based trust is the highest correlated with satisfaction (r=0.62) whereas emotion-based trust is the highest correlated with recommendation intention (r=0.51), which suggests that positively designed interfaces have the potential to be used to promote word-of-mouth marketing effectively. Nevertheless, individual differences exist. Some participants had made it clear

during interviews that they were against anthropomorphic features, indicating that the AI designs need to provide different degrees of personification.

Finally, the digital divide persists during intelligent transformation. Users of AI guides include mostly young people (the group between 18 and 30 makes up 58.0%), well-educated people (those with a degree of bachelor and above make up 71.6%), and technology-savvy users (users that often use AI products make up 67.0%), whereas those middle aged and older make up less than 15%. Intelligent transformation of technology does not necessarily lead to leveling of services but can lead to new forms of inequality in cultural participation.

The practical implications are represented in three aspects. Firstly, create a fault tolerant remediation system whereby if the AI is unable to answer or responds wrongly it should not suddenly refer the users to another source or provide wrong information but rather act truthfully, compensate and improve its performance.

Make the warmth available to AI, but provide an adjustment capability. On the one hand, add a memory preference option, including the storage of user visit paths or the use of personalized nicknames, in order to develop a feeling of respect and remembrance. Conversely, provide flexible degrees of anthropomorphism. Information accuracy is focused on in Professional Mode whereas Companionship Mode helps to express emotions better enabling users to select the way of interaction they like most.

In terms of management, make sure to incorporate the interactive experience into the main metrics of procurement assessments and effectiveness monitoring, and add the user testing stages. Also, it is advised to have a dual-track service policy. The preservation and optimization of conventional services, including human-assisted tours and paper-based guides, will be used to make sure that people who do not use digital technologies are not left out, and the gap in digital accessibility is reduced.

To sum up, on the industrial scale, the industry ought to promote a human-focused AI approach, which entails understanding that the purpose of AI is not to duplicate people; instead, it is to benefit them. Design of AI-based tours in the future must evolve along the lines of user requirements rather than the technological framework, taking into consideration the variety of tastes of users. Evaluation requirements of AI-friendly trust are encouraged to be set by industry including measures like fault tolerance and restoration of trust abilities with the aim of directing product development towards more stable reliability rather than pure accuracy.

5. Conclusion

The cyber guide at museums does not only represent the technological progress but also the transformation of human-machine interaction. The report indicates that trust in AI-guided tours by users is based not on technical approval alone, but a combination of functional experience, quality of interaction and emotional connection-users assess both the accuracy of the AI as well as its human touch. It is an object of academic study as well as a pre-condition of technology actually benefiting humanity. With the growing infiltration of artificial intelligence into the fabric of daily life, the question of why do humans trust machines is effectively the same as the question of what makes humans human. Maybe the answer can be found among these unexpected moments in the museum when the AI remembers a name of a user and also in the reaction of forgiveness of a user after an incorrect response: Oh well, it is only a machine. The future of human-machine relationships is not in the technology, but the way humans describe their interactions with it.

References

[1] Guo Quanzhong, Luo Jiale and Gu Kexin. AI agents as social actors: autonomy evolution, communication network embedding and risk reflection. News Enthusiasts, 1-17.

[2] Mayer, R.C. and Davis, J.H. (1995) An integrative model of organizational trust. Academy of management review, 20, 709-734.

[3] Licence. (2026). Privatisation du droit et limites comportementales des agents dAI. East Asian Law, (01), 15-28.

[4] Bach, Tita and Amna Khan and Harry Hallock and Gabriela Beltrão and Sonia Sousa. (2022) A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective. International Journal of Human-Computer Interaction. 40. 1-16.

[5] Hou, Pingping, Qiu, Hanqin, Ren, Lianping, Jia, Binyao, and Wang, Jingna. (2025). The study of trust of visitors to human-robot interaction in the museum environment: an elaboration likelihood perspective. Journal of Hospitality and Tourism Technology, 1 -16..

[6] Wooldridge, M., and Jennings, N. R. (1995). Intelligent agents : theory and practice. Knowledge Engineering Review, 10(2): 115-152.

[7] Jia Di. (2020). Design and Development of a Human-Computer Interaction System of Museum Tour Robots (Master thesis, Beijing Jiaotong University). The master degree.

[8] The West Lake Museum of Hangzhou has introduced the option of AI Image Recognition which allows its visitors to explore more about the history of cultural relics.

[9] National Cultural Heritage Administration. (2025) National Museum Directory of China in 2024. Beijing: National Cultural Heritage Administration, pp. 23-24.