Exploring Artists' and Art Viewers' Perspectives for Art Chatbots: Implications for a Design Framework

Jinyu Liu
Department of Computer Science
Durham University
Durham, United Kingdom
vvmj34@durham.ac.uk

Effie Lai-Chong Law
Department of Computer Science
Durham University
Durham, United Kingdom
lai-chong.law@durham.ac.uk

Hubert P. H. Shum
Department of Computer Science
Durham University
Durham, United Kingdom
hubert.shum@durham.ac.uk

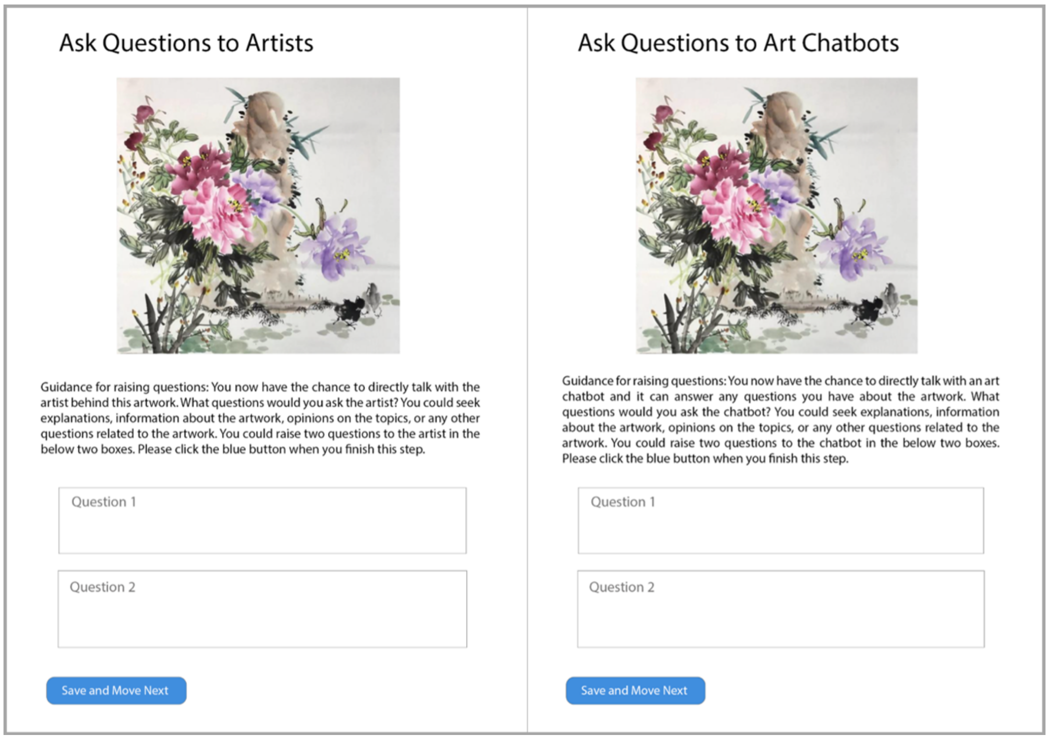

Figure 1: An example of interface comparison for the question part. The left interface is what participants see in the Artist condition and the right interface is in the Chatbot condition. Our participants see the interface in its Chinese version.

Abstract

Recent advances in large language models (LLMs) and conversational user interfaces (CUIs) unlock new ways to help art viewers get answers about artworks. To clarify the roles that artists and viewers envision for art chatbots, we conducted two empirical studies in the domain of traditional Chinese painting, given its cultural depth. First, we interviewed five artists about how they currently respond to viewer inquiries and their attitudes toward chatbots. Second, we asked art viewers (N=102) to pose questions to either an artist or a chatbot. Results show that artists see chatbots as useful for factual or repetitive queries but hesitate to entrust emotive or personal discussions to them. Viewers also favor chatbots for efficiency but desire human input for deeper or personal topics. Based on these insights, we propose a design framework that balances the perspectives of both artists and viewers, contributing to the CUI community's understanding of domain-specific chatbot design.

CUI '25, Waterloo, ON, Canada

ACM ISBN 979-8-4007-1527-3/2025/07
https://doi.org/10.1145/3719160.3736626

CCS Concepts

• **Human-centered Computing**; • **Human Computer Interaction (HCI)**;

Keywords

Art chatbots, Art appreciation, Art queries, Question-answering systems, Human-centered AI

ACM Reference Format:
Jinyu Liu, Effie Lai-Chong Law, and Hubert P. H. Shum. 2025. Exploring Artists' and Art Viewers' Perspectives for Art Chatbots: Implications for a Design Framework. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25), July 08–10, 2025, Waterloo, ON, Canada.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3719160.3736626

1 Introduction

Artwork serves as a medium for creating dialogues between artists and viewers [4]. However, many art viewers visiting galleries still leave with unanswered questions related to the artworks they viewed [32]. Although research shows dialogue fosters appreciation among viewers [34], the majority of artists may not have the time to respond to the viewers' questions [7]. Moreover, viewers may approach artworks in different ways: individual traits such as curiosity could affect both the kinds of questions they ask and the depth to which they pursue answers [20]. This gap points to the need for systems that efficiently respond to viewers' questions while offering possibilities for personal and in-depth discussions.

Art museums and galleries have experimented with chatbots for the delivery of information about artworks and exhibitions [25, 29]. However, integrating such technology effectively into the art context remains a challenge. There are issues related to the limitations of natural language processing [26] and unclear boundaries concerning the responsibilities of chatbots [41]. On such grounds, our research innovatively investigates the perspective of both artists and viewers to propose a design framework for art chatbots. By making clear the scope these chatbots can support artists in their practice and enhance viewers' experience, we position ourselves within the current debates around the application of chatbots in cultural contexts. Central to our inquiry are three research questions (RQs):

RQ1: To what extent do artists wish to delegate queries to a chatbot?
RQ2: To what extent are questions posed by art viewers to artists different from those to art chatbots in terms of question type and complexity?
RQ3: What implications do these preferences have for designing a chatbot that simultaneously meets the needs of viewers and artists?

To address these questions, we conducted two empirical studies. In the first study (Study 1), we involved five artists specializing in traditional Chinese paintings to explore their current practice in answering viewer queries, as well as their attitudes and expectations about a prospective art chatbot. Traditional Chinese painting, with its deeply embedded cultural and symbolic elements, naturally prompts different types of questions from viewers. Although Chinese art is attracting growing global interest, it remains relatively underexplored from a chatbot perspective, making it a fertile ground for examining how culturally nuanced inquiries could be addressed. Our results show that while the artists appreciate chatbots for handling repetitive questions, they are concerned about entrusting subjective or emotionally charged questions to a chatbot. In the second study (Study 2), we involved 102 art viewers from four different galleries, who, after viewing some artworks, were instructed to pose questions to either an art chatbot or an artist, enabling us to compare questioning attributes across the two conditions. This approach allowed us to identify key parallels and discrepancies between artists' and viewers' expectations for art chatbots. The viewers' questions were categorized into three main types: Artwork Fundamentals, Interpretive Perspectives, and Artistic Exploration (details see Section 5.1). Demographic variables were included as control variables to ensure their potentially confounding effects were accounted for; they were not central to the research questions of our work.

Synthesizing the insights from both studies, we observed that viewers' perspectives on the role of art chatbots align closely with those of artists in many aspects. Both groups favor chatbots for handling factual and repetitive questions, while reserving subjective or emotionally nuanced queries for artists. Building on these insights, we propose a framework to guide curators of galleries, designers, and researchers in creating LLM-powered art chatbots that accommodate the expectations of these stakeholders. In particular, we advocate for an application that handles factual queries while deferring more interpretive or personal questions to artists. Such a hybrid strategy maintains the authenticity of the artistic process and capitalizes on the chatbot's strengths in rapid information retrieval. This paper makes three key contributions:

- Insights into artists' perspectives on the potential role of art chatbots in their practice, particularly the types of questions they would delegate to such systems.
- An empirical study demonstrating a methodological approach to identifying art viewers' questioning patterns by comparing inquiries made to chatbots versus those directed at artists.
- A design framework that integrates these perspectives, providing guidelines for chatbot functionality, user experience, and policy considerations.

2 Background and related work

2.1 Promoting Art Viewer Engagement

Viewer engagement is significant for the appreciation of artworks [13]. Engaged viewers often develop a deeper understanding of artists' intentions and artworks' context, which can lead to more meaningful and memorable experiences [48]. This engagement fosters curiosity and dialogue, enabling viewers to connect emotionally and intellectually with the art [39]. Art viewers frequently seek experiences that allow them to explore the story behind artworks and gain deeper insights into artistic processes [28]. Many also expect opportunities for self-reflection through their experiences [31]. For many artists, such an exchange also serves as an enriching experience of gaining new insights into their work [2, 37].

Recent technology development has reframed such interactions, going beyond face-to-face discussions by including the option of posing digital comments on art platforms [41]. In the realm of Conversational User Interfaces (CUI), recent studies have explored how to design and build chatbots for cultural contexts [29], complemented by growing interest in the design of inclusive CUIs for domain-specific content [24]. This increased popularity of chatbots enables new ways of immediate, guided interaction which can connect viewers directly to information about the artwork. Major art institutions, such as Centre Pompidou, Pinacoteca de São Paulo, MoMA, and Akron Art Museum, have leveraged chatbots to provide information on artworks, exhibitions, and artists. These institutions have taken different approaches, with some focusing on delivering facts quickly and others prioritizing playful and interactive experiences [19, 30, 45, 46]. The goal of all these various implementations is to increase the engagement of viewers by having easier access to art-related data.

2.2 Questioning Behaviors and Answering Mechanisms in Art Appreciation

Questioning behavior has long been considered a central aspect of the art appreciation process [44], as it fosters active engagement and enhances the viewing experience [6]. For example, asking about the symbolism behind a specific color choice in a painting can reveal layers of meaning that might otherwise remain unnoticed. Traditionally, these interactions were mainly realized through art tours or talks organized by curators or artists, which had obvious limitations for scalability. As galleries expand into virtual offerings, automated QA systems like chatbots have created new possibilities for handling large volumes of queries [25]. The solutions offered by chatbots include rule-based and generative AI (GenAI) approaches. Rule-based chatbots rely on handcrafted scripts, offering precise and controlled answers to frequently asked questions, but struggle with open-domain scalability and complex queries [54]. GenAI-based chatbots enabled by LLMs provide better generalization and handle diverse, open-ended questions; however, they risk generating hallucinations [55].

2.2.1 Measuring Question Complexity. In order to design effective question-answering mechanisms, it is important to consider the varying levels of complexity in the questions themselves. Whether viewers ask about straightforward details or seek deeper insights, each inquiry requires a different degree of cognitive processing. Analyzing these differences can guide chatbots in providing responses that appropriately match a viewer's needs. The Webb's Depth of Knowledge (DOK) framework has been widely adopted to measure the cognitive complexity of questions and tasks in education [11]. It categorizes tasks into four levels: Level 1 involves recall and reproduction of facts, Level 2 focuses on skills and concepts, Level 3 requires strategic thinking and reasoning, and Level 4 emphasizes extended critical thinking and synthesis. This framework helps assess and design interactions that align with the varying complexities of art inquiries. Originally developed for K–12 assessments, DOK has been integrated into arts education, aligning analytical or interpretive tasks with DOK levels [27]. By designing responses according to question complexity, educators can foster engagement, moving beyond surface-level queries toward deeper analysis [16]. As it demonstrates suitability to measure the complexity of art inquiries, in this study, we apply the DOK framework to categorize and analyze the complexity of questions posed by art viewers. By mapping each question to one of DOK's four levels ranging from basic recall of factual details to advanced critical thinking, we provide a structured understanding of how question complexity varies between artists and chatbots.

2.2.2 Influences on Questioning Behaviors. Some personal attributes can impact the viewer's questioning patterns. For example, curiosity has been linked with the motivation to achieve new knowledge by asking questions at a deeper level [20]. In an art context, viewers with higher curiosity may ask for more details beyond superficial information. When interacting with chatbots, these viewers may also stretch the limits of chatbots to request more personal insights. In addition to curiosity, demographic factors and artwork styles may also influence questioning behavior. For instance, gender may shape the way individuals approach inquiries [12], while the painting theme, such as Landscape Painting in Chinese Art, could elicit different types of questions based on their inherent characteristics [10].

Previous research suggests curiosity level can be measured by the Curiosity and Exploration Inventory-II (CEI-II) [22]. Understanding these personality-driven motivations will help in designing chatbots that could adapt their responses based on different dimensions in viewers' questions like complexity. In our study, we use CEI-II to measure the curiosity levels of art viewers, helping to understand how this personality trait influences their questioning behavior. Specifically, CEI-II provides a structured framework to measure curiosity, which is then correlated with the complexity and types of questions posed by viewers. Additionally, we examine whether demographic factors like gender and painting themes correlate with question types or complexity, providing a more comprehensive view of how these attributes shape questioning patterns.

2.3 Chatbots in Art and LLM Integration

2.3.1 LLM Integration. Technology has played a central role in the development of how viewers interact with art [8, 17]. Large language models (LLMs) are capable of processing complex questions and giving real-time responses [1]. For instance, in an art context, an LLM can analyze a viewer's question about the historical influences on a painting's style and provide a concise summary of relevant art movements. This adaptability makes them attractive for handling the rich diversity of viewer queries in art settings. However, LLMs may also produce incorrect or fabricated information referred to as "hallucinations" [33], which could pose risks for art institutions seeking to maintain credible, authentic narratives. Determining the question types that can be safely answered by an LLM-based chatbot versus those best forwarded to an artist remains a crucial challenge. Furthermore, researchers increasingly recognize that chatbots can influence user behavior [18] and users may overestimate chatbot capabilities [23]. These concerns highlight the importance of disclaimers and well-defined boundaries to prevent misinformation [14]. Moving from these considerations about LLM capabilities and pitfalls, it is essential to address how to weigh the trustworthiness of chatbots compared to human experts.

2.3.2 Trust with Human Experts vs. Chatbots. While many LLM-based chatbots demonstrate abilities to handle general questions more effectively than human beings [36], people still seem to trust human experts for certain types of questions [3, 15], especially in contexts that call for deeper knowledge or insight into nuances. Previous research suggests that humans tend to regard chatbots as a tool for factual retrieval rather than a source of interpretive wisdom [9]. Studies also indicate that while chatbots are perceived as friendly and convenient, credibility remains more strongly associated with human professionals [3]. However, some users do treat

Table 1: Details of the participants in the Artist Study

User	Gender	Age	Education	Occupation	Art Experience(Years)	Number of Past Exhibitions
A1	M	33	Master of Fine Arts	Painter, Art College Instructor	10	9
A2	F	45	Bachelor of Arts	Curator, Painter, Photographer	25	13
A3	M	39	Doctor of Art History	Painter, University Lecturer	17	12
A4	F	52	Bachelor of Art History	Art Journal Writer, Painter	28	17
A5	F	31	Master of Arts	Designer, Painter, Gallery Manager	11	8

chatbots almost like human conversational partners, particularly if they have experience interacting with advanced AI [35]. Recent investigations in the CUI area point to "persona design" as a promising strategy for bridging the gap between factual responses and more affective dialogues [40]. This discrepancy underscores the importance of designing chatbots that satisfy users seeking deeper engagement.

3 Study 1 - Artists

3.1 Methods

3.1.1 Participants. We conducted semi-structured interviews with five artists in June and July 2024 to discuss how they envisioned interacting with art viewers, how they respond to their questions, and what challenges or opportunities might arise from art chatbot integration. They were recruited through the personal networks of the first author, who is an artist specialised in traditional Chinese painting. While these artists hold various roles, such as curators and art teachers, all of them are painters with expertise in traditional Chinese painting. This shared background ensured consistency in the artistic focus of Study 2 with art viewers. We selected these artists based on their educational and curatorial backgrounds, with extensive experience in teaching at schools or collaborating with galleries. Participation was voluntary. Table 1 provides detailed information about the artists.

3.1.2 Procedure. Before conducting the interviews, we checked with participants about their understanding of chatbots to ensure they had a clear definition of what a chatbot is. Each following interview lasted about 40 minutes and comprised three parts where the artists were asked to

- Describe how they generally communicate with viewers and how they manage follow-up conversations.
- Reflect on difficulties they encounter in responding to viewer inquiries.
- Discuss ways they believed an art chatbot could help or influence their interactions with viewers.

All interviews were audio-recorded, and anonymized transcripts were produced. Three interviews were conducted online via Zoom, and two were conducted in person at their university offices where the artists work. These settings ensured a quiet environment without distractions, conducive to open discussions. We then conducted a thematic analysis [21], refining themes through multiple reviews and coding quotes to illustrate recurring ideas (Section 3.2). The first coder, who has expertise in art and qualitative analysis, performed open coding on all transcripts to develop an initial set of themes. The second coder, who has extensive interview research experience, independently reviewed 20% of the coded excerpts. Initial agreement was 78% across coded excerpts. The two coders held rounds of discussions to resolve discrepancies and refine the codebook until full agreement was reached. Here we illustrate with two cases of disagreement:(1)A quotation from A3—"I rarely have time to talk after an opening; I rush to the next activity"—was labelled by the first coder as Time Constraints but by the second as Viewer-Engagement Barriers.(2)For A5's remark, "When people ask about the emotional journey behind my painting, I believe they are seeking a personal connection; I think it's hard for a chatbot to effectively convey these intimate stories," one coder tagged Need for Human Touch, while the other chose Chatbot Limitations. After discussions, the two coders converged on Time Constraints on the first case and Need for Human Touch on the second case. Codes were then clustered into higher-level themes, which are reported in Section 3.2. To ensure the accuracy of the transcripts, they were translated from Chinese to English by experienced translators with an artistic background. Additionally, we performed back-translation by randomly selecting a subset of the translated transcripts and translating them back into Chinese to verify consistency and accuracy. This process ensured that technical art terms and nuanced expressions were correctly interpreted, preserving the integrity of the original content.

3.2 Findings from artists (Study 1)

3.2.1 Engagement with artists. In the interviews, five artist participants (A1 to A5) talked about some common methods to answer questions from art viewers. Most communication happens through face-to-face interactions during public art talks and workshops (A1-A5). Some artists also prefer answering questions through online platforms in recent years (A1, A5). All artists thought it was useful to hear questions from viewers (A1-A5). However, artists mentioned that having enough time to answer viewers' questions was their biggest challenge (A1-A5). They often had to prepare new artworks or exhibitions (A2-A4). Some artists also had other work duties like teaching (A1, A3). These difficulties made it hard for them to handle viewers' questions. Even when they had time in some art talks, they could not respond to questions in detail (A4, A5). This was because they often had to prepare presentation materials or new works on short notice, and balancing deep follow-up discussions with a busy event schedule and a large number of viewers made it impractical to address every query in-depth.

Participants mentioned most of the galleries lacked question-answering mechanisms that viewers could effectively leave their questions and receive feedback (A3-A5). Some galleries might not have sufficient staffing to deal with viewers' inquiries (A1, A2). Some larger galleries invited viewers to artist talks for communication, but these chances were rare and often only opened to limited members (A3). When discussing the use of chatbots, A1 and A5 mentioned that they noticed that some galleries have experimented with chatbots, but these chatbots mainly deal with common queries and cannot answer complicated art-related questions.

3.2.2 Integration of art chatbots. Before discussing specific functionalities of future art chatbots, participants shared their interpretations of what chatbots are and their previous encounters with chatbots. Although the researchers provided a standardized definition of chatbot prior to interviews, artists interpreted the concept based on their own experiences. A1 viewed chatbots as tools designed to answer questions and provide instant feedback. Earlier, he used ChatGPT to learn about the history of certain forms of art and to develop creative ideas. A3 defined chatbots more narrowly as an automated system for answering frequently asked questions. He reported familiarity with their implementation within the art domain and thought their functionalities remained no more than answering general-level questions. A5 thought that a chatbot is generally a digital agent that can help "sort questions and direct people further to relevant places.". While A2 and A4 never used chatbots personally, they both knew of their increasing utilization to provide responses and were curious about the use of such technological applications to handle viewer questions. We then asked participants what features of chatbots they thought would be useful in the art domain. A common function mentioned was providing support to deal with informative and repetitive questions (A1, A2, A4). A3 hoped that chatbots could organize viewers' inquiries into categories and forward the questions to him.

Participants talked about some potential challenges they can imagine with chatbots. They were concerned about the ability of chatbots to provide satisfactory answers to questions related to personal and subjective experience (A2-A5). For example,

> *"The mood conveyed in a painting can carry a lot of nuanced information. I typically tailor my response based on who is asking the question, rather than having a fixed answer. That's why it's challenging for me to provide all the background details to a chatbot." (A2)*

> *"When people ask about the emotional journey behind my painting, I believe they are seeking a personal connection. I think it's hard for a chatbot to effectively convey these intimate stories." (A5)*

Besides, based on the description from A3, the reason for people to come to art talks is "*not just to see the artwork but also wish to interact directly*". In addition, several artists (A1, A3, A5) suggested that periodically interviewing artists to refine the chatbot's knowledge base or functionalities would be useful. They believed this approach would ensure the system to be kept up to date with artists' evolving perspectives, helping the chatbot reflect the changing intent behind new artworks and maintain accurate content. All five artists welcomed the idea of receiving aggregated viewer-question reports, noting that such insights could "spark new creative directions" (A3) and "help us tailor future talks" (A1). They also requested the ability to screen any public-facing responses to safeguard personal narratives.

4 Study 2 – Viewers

Following Study 1 with artists, we present Study 2 with art viewers, detailing the respective process of data collection. The two studies were both granted ethics approval by the Ethics Committee of Durham University.

4.1 Methods

4.1.1 Participants. In Study 2, we invited a total of 106 viewers from four galleries in China (around 26 viewers per gallery) to participate. Galleries 1 and 2 are art studios with gallery functions owned by the artists themselves, showcasing artwork, with a special focus on traditional Chinese paintings. These galleries are approximately 50 sq meters in size, hosting around 10 to 20 viewers each day. Galleries 3 and 4 are small private galleries in communities that focus on general art exhibitions. Each is about 100 sq meters in size, drawing approximately 20 to 50 viewers daily. We excluded 4 participants for incomplete responses, leaving a final sample of 102 participants. This sample size ensures sufficient data for question-type and complexity analysis (Section 4.1.3). Each participant was randomly assigned to either the Artist condition or the Chatbot condition (51 in each condition), following the between-subjects experimental design. Study 2 was conducted in August and September 2024. Participants represented diverse demographic backgrounds (see Figure 2), which helped us capture varied questioning behaviors. Among the 102 participants, ages ranged from 18 to 84 years (M=39.5, SD=16.1), with 47% male and 53% female participants. The distribution of Educational Level was 32% University level, 15% Postgraduate level, and 53% Pre-University Level. The participation of all art viewers was voluntary. It took around 10 minutes for each participant to complete the tasks (see Section 4.1.3).

4.1.2 Artwork Selection. To prepare Study 2 on comparing questions posed to artists and chatbots, we asked each of the five interviewed artists to provide ten pieces of traditional Chinese painting from their past exhibitions, for a total of fifty candidate artworks. Traditional Chinese painting was chosen for its cultural resonance with both Chinese artists and the local Chinese viewers in our study, offering varied opportunities to spark questions about symbolism and interpretive meaning. To limit the scope while maintaining variety, we finally selected twelve paintings across three themes (four works per theme), including Bird and Flower Painting, Landscape Painting, and Figure Painting. These three themes (see Figure 3) were considered as they represent the main pillars of traditional Chinese painting [38, 43], together representing the diversity and richness of this traditional art form [42]. When choosing these themes, we considered their potential to elicit a broad spectrum of viewer responses. Bird and Flower Painting carries natural symbolism that invites more interpretive questions, while Landscape Painting evokes more thematic and contemplative questions in its philosophical underpinnings. Figure Painting, with its emphasis on human narrative and emotion, tends to yield storytelling and identity questions. These inherent differences in visual and symbolic content were consciously taken into consideration with the aim of

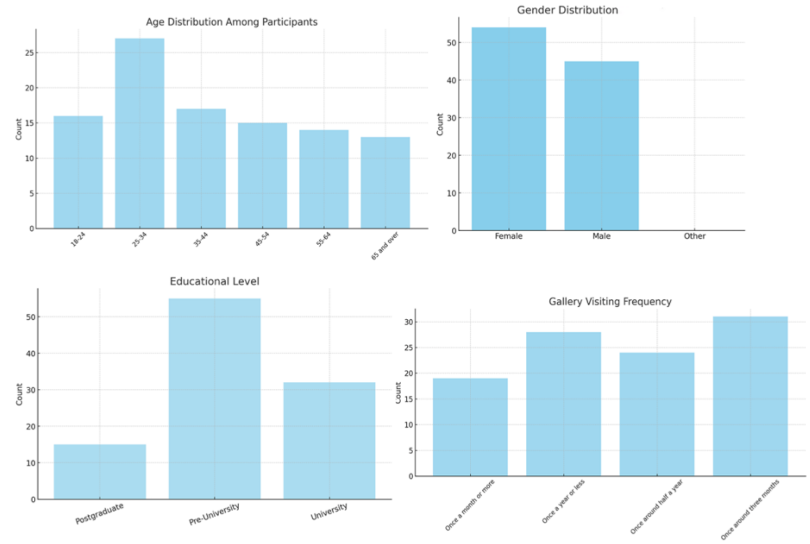

Figure 2: Demographic information of viewer participants (Age/Gender/Educational Level/Gallery Visiting Frequency).

Figure 3: Selected traditional Chinese paintings representing the three themes used in Study 2: Landscape Painting (left), Bird and Flower Painting (middle), and Figure Painting (right).

ensuring a balanced basis for comparison. By choosing different and complementary themes, we tried to reflect the breadth of traditional Chinese painting. Artist-identifying details were removed to avoid bias. Each participant was randomly assigned three paintings (one from each theme), ensuring exposure to a wide range of themes.

4.1.3 Empirical Study Design. For Study 2, we developed a tablet-based application for each participant to use through several steps:

1. **Study overview**: As a preparation before viewing the artworks and submitting their questions, each participant needs to read the participant information sheet and then give their informed consent to take part in the study. Before proceeding, we also verified participants' understanding of chatbots to ensure they had a clear and accurate definition.

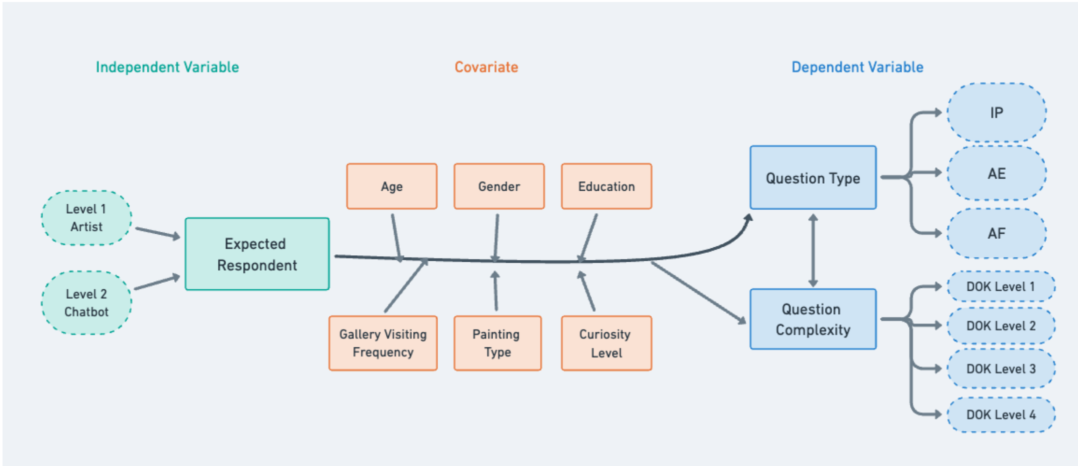

Figure 4: The research model of Study 2, showing how the main variables (Expected Respondent, Covariates, and Dependent Variables) interrelate. Note: IP = Interpretative Perspectives, AE = Artistic Exploration, AF = Artwork Fundamentals, DOK = Webb's Depth of Knowledge framework.

2. **Artwork viewing and question submission** (Figure 1): All artworks were displayed as high-resolution images on the tablet application; participants did not view physical paintings, ensuring consistency across sessions. Each participant was shown one painting at a time. Depending on their assigned condition (Artist or Chatbot), they were instructed to pose two questions to either an artist or an art chatbot. They were also informed that responses would not be immediately provided. This procedure was repeated for all three artworks, yielding six questions per participant.
3. **Curiosity questionnaire and demographic survey**: At the end of this experiment, all viewers completed the Curiosity and Exploration Inventory-II (CEI-II) survey to measure their curiosity levels (see Section 2.2.2 and Appendix A) and a demographic survey (Figure 2).

Figure 4 illustrates the research model for Study 2. The between-subjects approach (artist vs. chatbot) minimized learning effects that often arise in within-subjects designs, where participants become familiar with the task. By assigning participants to only one condition, we could more directly observe how they posed questions to different intended respondents (artist vs. chatbot). Moreover, gathering questions in a scenario where no immediate feedback was provided helped isolate questioning behavior from the influence of artists' or chatbots' actual responses, which can shape follow-up queries. Although this arrangement implied that participants did not receive answers, it allowed a clearer view of the initial questions viewers are likely to ask. Participants assigned to the Chatbot condition were informed that their conversational partner was an AI-powered system built on a large language model, not a human artist. No measures were taken to adjust participants' preconceived notions about AI capabilities; thus, their expectations regarding AI may have influenced the questions they posed.

We recognized that chatbots may exhibit limitations such as producing inaccurate or biased text. However, as our participants only generated questions without receiving real-time feedback, these limitations did not directly affect our data collection. The study focused on what viewers ask, rather than how chatbots or artists respond. Although the scenario of posing questions without immediate answers is somewhat artificial, logistical constraints (e.g., real-time artist availability) and the risk of LLM "hallucinations" made this design a pragmatic choice. Future studies may incorporate real-time responses to explore how dynamic interaction influences follow-up questioning.

4.1.4 Instruments. To measure participants' curiosity levels, Curiosity and Exploration Inventory-II (CEI-II) was used in our study (Section 2.2.2). The inventory includes ten items such as "I actively seek out new experiences" and "I am deeply immersed in the activities I enjoy." (See Appendix A for the full list of items). Each item is rated on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). The total curiosity score is calculated by summing the ratings across all items, with higher scores indicating greater curiosity, and the range is 10 to 50. These scores were analyzed to understand their relationship with question types and complexity in this study.

Webb's Depth of Knowledge (DOK) framework was employed to assess the complexity of questions posed by participants (Section 2.2.1). This framework categorizes questions into four levels based on their complexity attributes.

- **Level 1:** Recall and reproduction
- **Level 2:** Skills and concepts
- **Level 3:** Strategic thinking and reasoning
- **Level 4:** Extended critical thinking and synthesis

The inventory ensured a consistent classification of questions across these four levels. We used the DOK criteria, adapted for art appreciation contexts (see Appendix B), to categorize and analyze the complexity of questions posed to artists and chatbots.

4.1.5 Data Analysis Methods. We adopted a multi-faceted approach in analyzing the questions collected from viewers, including thematic analysis and complexity analysis with relevant frameworks and tools. Each analysis is detailed in the following. We first conducted thematic analysis to identify patterns in question categories. The first researcher, who has a systematic art training background and 6-year experience with data analysis, initially coded all questions and generated a preliminary codebook detailing clear definitions and examples for each category and sub-category. The second researcher, who is also an artist with painting expertise and has eight years of experience in qualitative analysis, applied this codebook to a 30% random sample of collected questions, achieving a Cohen's kappa of 0.85, indicating strong inter-rater reliability. All discrepancies were then discussed and resolved to refine category definitions as needed. With the codebook validated, the first researcher proceeded to code the full dataset.

Following this, we analyzed the complexity of the questions with the use of Webb's DOK framework (see Section 2.2.1). The criteria at each level of DOK were adapted to an art appreciation context. For example, Level 1 questions deal with recalling factual details (e.g., "What material is used for this painting?"), while Level 4 questions call for synthesis or critical evaluation of information (e.g., "How does the color choice of this piece reflect societal changes during its creation time?"). Both researchers independently classified an initial subset of 30% of questions, yielding a Cohen's kappa of 0.743, which falls within the substantial agreement range (0.61-0.80) according to Landis and Koch [53]. Although Cohen's kappa indicates substantial agreement, some discrepancies remained. A typical disagreement illustrated why discussion was needed: one viewer asked, "Why did the artist choose a vertical scroll instead of a horizontal one?"—the first coder marked this as Level2 (Skills & Concepts), whereas the second coder argued for Level3 because it required reasoning about compositional intent. After discussions, the two coders converged to mark this as Level 2. The coders reviewed each discrepant item together, identified the source of disagreement, and re-examined the questions against the DOK framework criteria. Through this discussion, the codebook definitions and examples were further updated before being applied to the full dataset. This approach allowed a clear comparison of cognitive complexity between the questions being posed to artists versus chatbots.

5 Findings from Art Viewers (Study 2)

Following the data analysis methods described in Section 4.1.5, we categorized and interpreted the questions posed by viewers under both Artist and Chatbot conditions. To provide deeper insights, this section is structured into two main analyses: Question Category Comparison and Question Complexity Analysis. Each analysis starts with qualitative findings followed by quantitative results.

5.1 Question Category Comparison

5.1.1 Qualitative Findings. To understand how viewers approach different dimensions of art inquiry, we conducted a thematic analysis focusing on the content and focus of their questions. The findings revealed three main categories:

- **Artwork Fundamentals (AF)**: Questions seeking factual details about the artwork and clarifying why and how the artist made specific choices. This category includes sub-categories such as Core Facts (e.g., "When was this painting made?").
- **Interpretive Perspectives (IP)**: Questions exploring the deeper meaning of the work, focusing on symbolism, emotion, and the artist's influences or philosophy. Sub-categories include such as Emotional Tone (e.g., "What feeling did you want to convey?").
- **Artistic Exploration (AE)**: Questions relating to the creative process and appreciation of art, offering guidance on the artist's workflow, techniques, and ways to develop one's own skills and understanding. Sub-categories include such as Appreciation Methods (e.g., "How should one approach interpreting abstract shapes?").

Table 2 provides an overview of the three main categories, their sub-categories, and examples. From the qualitative results, we found that viewers tended to ask more factual and process-oriented questions (e.g., Core Facts, Creative Details) when they expected to interact with a chatbot, while interpretive and reflective queries (e.g., Emotional Tone, Artist Reflection) were more common when they expected a human artist. This divergence may stem from the perception of chatbots as efficient tools for retrieving objective information, whereas human artists are often viewed as the primary source for conveying nuanced and emotional insights. Prior studies have also suggested that social presence and authenticity play a significant role in shaping expectations, with users tending to reserve deeper, interpretive inquiries for contexts involving human expertise [5, 47].

5.1.2 Quantitative Result Analysis. To quantitatively assess which factors most strongly predict the type of question, we conducted a multinomial logistic regression to investigate how different demographic and contextual factors (Gender, Age, Educational Level, Gallery Visiting Frequency, Painting Theme, Curiosity Level and Expected Respondent) predict the type of question asked (AF, IP and AE) (cf. the Research Model in Figure 3). This approach was chosen because our dependent variable, Question Type, consists of more than two non-ordinal categories.

The final model fits significantly better than an intercept-only model (χ^2 (22) = 42.46, p = 0.005). The Pearson's test suggests that the model is a good fit for the data (Pearson χ^2 (588) = 567.540, p = 0.720). Table 3 presents the results from Likelihood Ratio Tests for each predictor. As shown in Table 3, except for the predictor of Expected Respondent (i.e. Artist vs. Chatbot), all the other factors (Gender, Age, Educational Level, Gallery Visiting Frequency, Painting Type and Curiosity Score) were not statistically significant predictors of Question Type (all p > 0.05) (see Table 7 and Table 8 in Appendix 3 for details on the coefficient (B), p-values, odds ratios (OR) and corresponding 95% confidence intervals for each predictor). For Expected Respondent as a predictor variable, compared to the reference category (AF), Expected Respondent in the Artist Condition was associated with significantly higher odds of IP questions (OR = 2.85, 95% CI: 1.95-4.16, p < 0.001). The multinomial regression analysis indicates that Expected Respondent is the sole factor influencing Question Types posed by art viewers in our study. Other demographic or situational factors did not significantly predict differences in question types in this study.

Table 2: The question taxonomy from our qualitative analysis. We identified three main categories and ten sub-categories through the analysis.

Main Category*	Sub-Category*	Description	Example
Artwork Fundamentals (328)	Core Facts (67)	Focuses on purely factual details, like the artwork's dimensions, materials, date of creation, or other verifiable data.	What is the actual size of this painting?
	Creative Details (114)	Examines the artist's specific choices—why particular elements, colors, or subjects were selected and how they shape the artwork's concept.	Why did you choose such color for the flowers in this piece?
	Craft Techniques (115)	Covers how the artwork was produced, describing the methods, processes, tools, or technical steps employed in its creation.	How did you achieve the layered effect in this painting?
	Definite Answers (32)	Addresses yes/no or straightforward clarifications about the artwork, such as whether a certain feature or inspiration is part of it.	Is this painting based on a real location?
Interpretive Perspectives (213)	Emotional Tone (47)	Highlights the feelings or mood conveyed by the artwork, influencing the viewer's emotional response.	Are you trying to convey a feeling of loneliness in this painting?
	Symbolic Meaning (93)	Explores underlying ideas, symbolism, or references that add depth and interpretive layers to the piece.	Does the red color of the flower represent something specific?
	Artist Reflection (73)	Centers on the artist's personal perspective or philosophy, discussing their influences, motivations, or intentions behind the work.	How do you view the relationship between art and emotion in your creations?
Artistic Exploration (79)	Artistic Workflow (24)	Breaks down the sequence of steps the artist took to develop the piece, from concept sketches to final execution.	What was the approximate order when you created the elements in this painting?
	Skill Pathways (41)	Offers guidance for viewers on replicating or adapting the artwork's techniques, suggesting resources, practice methods, and best practices.	If I want to learn how to paint peonies, what tools and materials do I need to prepare?
	Appreciation Methods (14)	Suggests ways to view, interpret, or engage with the artwork, providing tips for deeper understanding and richer aesthetic experiences.	Is there a particular detail in this painting that you think viewers often miss?

*The numbers in brackets represent the question number for each category

Table 3: Likelihood Ratio Test Results

Predictor	-2 Log Likelihood	Chi-square	df	p-values
Expected Respondent	**920.8**	**30.4**	**2**	**<0.001**
Gender	891.5	1.04	2	0.594
Age	892.0	1.56	2	0.457
Educational Level	892.7	2.33	4	0.674
Gallery Visiting Frequency	895.3	4.86	6	0.562
Painting Theme	894.1	3.65	4	0.455
Curiosity Score	891.8	1.39	2	0.499

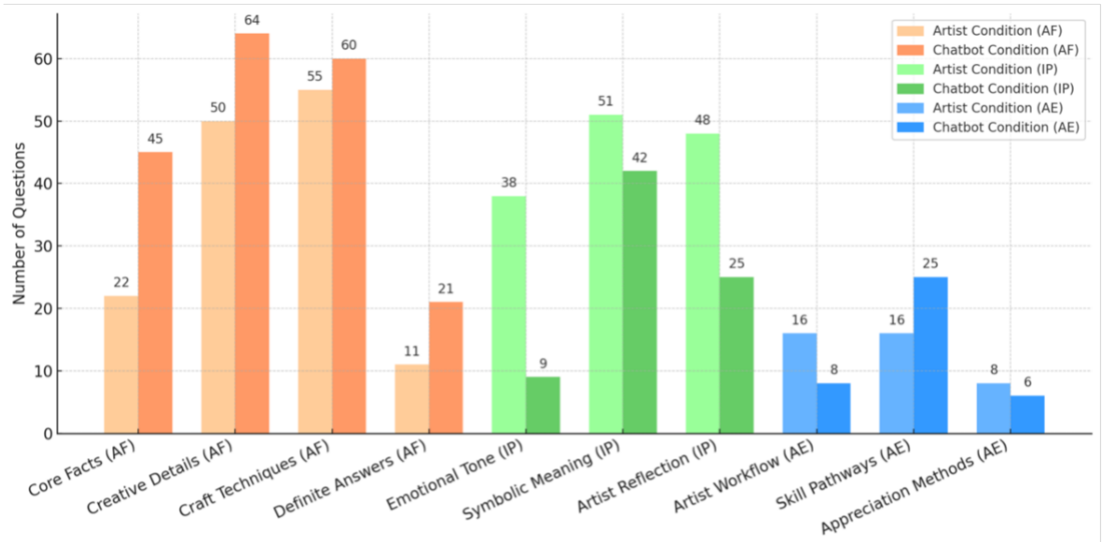

Figure 5: Distribution of questions by sub-categories to artists and chatbots.

Table 4: Number of questions in each question category for different question respondents

Condition	Artwork Fundamentals (AF)	Interpretive Perspective (IP)	Artistic Exploration (AE)
Artist	138	137	40
Chatbot	190	76	39

After establishing the overall model fit through the multinomial logistic regression, we further examined the distribution of question categories using the Chi-square test to confirm any significant association between Expected Respondent and Question Type (Table 4). The results show a statistically significant association between the two variables ($\chi^2 = 25.57$, $p < 0.001$). As shown in Table 4, the Chatbot condition received significantly more questions under the AF category. Within the Artwork Fundamentals category, sub-categories such as Core Facts and Creative Details were especially prominent for chatbot-directed inquiries (see Figure 5). However, viewers posed more questions in the IP category in the Artist condition, especially about Emotional Tone and Artist Reflection. Although the questions regarding Symbolic Meaning were relatively even, the general trend of addressing more interpretive topics to artists was nonetheless quite clear. In the AE category, questions were almost evenly split between the two conditions. Questions directed to artists primarily focused on Artistic Workflow and Appreciation Methods, while those aimed at chatbots centered somewhat more on Skill Pathways. Notably, viewers seemed to seek broader or multidimensional insights from artists, as evidenced by a higher frequency of questions bridging multiple categories (10 for Artist vs. 3 for Chatbot). These findings correspond to the qualitative results and underscore the observation that participants expect a richer emotional and conceptual dialogue with artists in general, while they turn to chatbots for more straightforward and factual information.

5.2 Question Complexity Analysis

5.2.1 Qualitative Findings. We employed Webb's DOK framework (Section 2.2.1) to categorize each viewer question according to its cognitive complexity—ranging from simple factual inquiries to more in-depth interpretive or analytical prompts. This approach helped clarify how complexity varied between the Artist and Chatbot conditions. As 5 shows, viewers asked artists proportionally more complex questions (DOK Level 3 and 4), reserving simpler, factual, or procedural inquiries (DOK Levels 1 and 2) primarily for chatbots.

In the Artist condition, for instance, participants posed interpretive questions such as "What is your motivation to use this particular red color palette for the facial expression of this character?" (DOK Level 3). By contrast, in the Chatbot condition, they tended to submit more lower-level factual queries like "Which location is this painting depicted?" (DOK Level 1). This pattern aligns with the artists' expectations—previously noted in Study 1—regarding personal engagement and deeper insights from human experts. As we discuss later in Section 5.2.2, these qualitative observations also correspond with the quantitative findings, further illustrating that question complexity and the expected respondent are closely linked.

5.2.2 Quantitative Result Analysis. An ordinal regression analysis was conducted to examine the relationships between the demographic and contextual factors (Gender, Age, Educational Level,

Table 5: DOK results for different expected respondents (Artist and Chatbot) with the question number and examples.

	Artist	Chatbot	Example
DOK 1	48	85	Which location is this painting depicted?
DOK 2	88	97	How did you create the texture of the cliffs?
DOK 3	162	123	Why did you choose to depict the trees on the mountains with such delicate details?
DOK 4	8	1	Do you view the theme in this painting as more of an exploration of nature, or a personal expression? How does this influence your brush control during the creation?

*The numbers in the table represent the number of questions for each DOK level.

Table 6: Ordinal Logistic Regression Results for Question Complexity

Predictor*	Coefficient	p-values	OR (95% CI)
Expected Respondent=Artist	**0.667**	**<0.001**	**1.95 [1.42, 2.67]**
Gender=Female	-0.282	0.091	0.75 [0.54, 1.05]
Age	0.001	0.903	1.00 [0.99, 1.01]
Education=Postgraduate	0.280	0.265	1.32 [0.81, 2.17]
Education=Pre-university	0.080	0.659	1.08 [0.76, 1.55]
Gallery Visiting Frequency = Once a month or more	0.222	0.328	1.25 [0.80, 1.95]
Gallery Visiting Frequency = Once a year or less	0.184	0.374	1.20 [0.80, 1.80]
Gallery Visiting Frequency = Once around half a year	0.338	0.114	1.40 [0.92, 2.13]
Painting Type = Bird and flower painting	-0.327	0.081	0.72 [0.50, 1.04]
Painting Type = Figure painting	0.244	0.201	1.28 [0.50, 1.04]
Curiosity Score	0.034	0.188	1.03 [0.98, 1.09]

*The reference categories for the ordinal logistic regression model: Expected Respondent = Chatbot, Gender = Male, Education = University, Gallery Visiting Frequency = Once around three months, Painting Type = Landscape painting

Gallery Visiting Frequency, Painting Theme, Curiosity Score, Expected Respondent) (cf. the Research Model in Figure 3) and Question Complexity, which was measured on an ordinal scale (DOK level 1 through 4).

The final model was a significant improvement over the intercept-only model (χ^2 (11) = 37.2, p < 0.001). The Pearson's test suggests that the model is a good fit for the data (Pearson χ^2 = 557.652, p = 0.886). Table 6 presents the coefficients, odds ratios (OR), confidence intervals (CI), and p-values for each independent variable. As shown in Table 6, only Expected Respondent reached statistical significance at the 0.05 level. Specifically, the log-odds estimate of Expected Respondent in the Artist condition was 0.667 (p < 0.001), corresponding to an OR of 1.95 (95% CI: [1.42, 2.67]). This means that participants expecting to pose a question to an artist had nearly twice the odds of asking a higher-complexity question than those expecting to interact with a chatbot. All other predictors (Age, Gender, Education Level, Gallery Visiting Frequency, Painting Theme, and Curiosity Level) were not statistically significant (all p > 0.05). Overall, the ordinal regression results suggest that Expected Respondent is the main factor influencing Question Complexity which aligns well with the qualitative observation.

After establishing the overall model fit through the ordinal logistic regression, we further examined the distribution of question complexity using a Chi-square test to confirm any significant association between Expected Respondent and Question Complexity. The test revealed a strong association between the two variables (χ^2 = 27.928, p < 0.001). This result highlights the tendency of art viewers to adapt the complexity of their questions based on the type of expected respondent. As shown in Figure 6, viewers posed significantly more complex questions (DOK Levels 3 and 4) to human artists, while chatbots predominantly received simpler and factual inquiries (DOK Levels 1 and 2). These findings emphasize the critical role of Expected Respondent in shaping the complexity of questions.

To further examine whether Question Complexity is associated with Question Types, we ran a Chi-square test comparing DOK level (1–4) and Question Type (AF, IP, and AE). Results showed a significant association (χ^2 (6) = 310.517, p < 0.001). From the test results, questions in the AF category predominate at DOK 1 and 2 which indicates lower complexity. Questions in the IP category are far more likely to appear at DOK 3 and 4, suggesting that interpretive topics often require more complex thinking. Questions in the AE category mostly fall into DOK 2, with fewer in DOK 1 or the higher DOK levels. This suggests that while queries about artistic exploration can require some conceptual thinking, they typically do not rise to the deep interpretive complexity.

6 DISCUSSION

6.1 Revisiting Research Questions

In this subsection, we revisit the three RQs (Introduction) by drawing on the empirical findings.

- **RQ1**: The five artists confirmed they appreciate having chatbot support for repetitive or factual queries (RQ1), but they

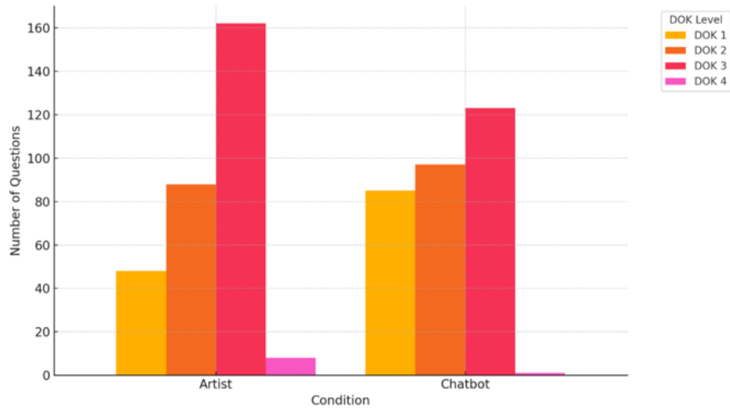

Figure 6: Complexity of Questions by DOK Levels between the two conditions.

were cautious about entrusting more subjective or emotional questions to chatbots (Section 3).

- **RQ2**: The data indicate clearly that art viewers significantly alter their questioning behaviors, depending on whether an artist or a chatbot is expected to answer a query (Section 5). Specifically, they commonly reserved interpretive and emotional questions for artists, leaving simpler and factual inquiries to chatbots. This observation aligns with previous studies about the social presence of conversational agents, where chatbots are primarily perceived more as information tools rather than empathetic conversational partners [5, 47].
- **RQ3**: Despite this broad trend, our findings also indicated that a subset of viewers are open to using chatbots for more subjective or emotional topics. It highlights a diversity of attitudes that could expand the chatbot's responsibilities if designed appropriately. In turn, these preferences imply that any implementation of art chatbots should maintain a careful balance: providing fast and accurate responses without influencing the personal connection that many viewers seek directly from the artist. Consequently, these insights show how chatbots can be used most effectively while preserving the authenticity that artists value. We further elaborate on these implications in Section 6.2, where we propose a design framework to address these considerations.

6.2 A Design Framework for Art Chatbot Design

Based on the empirical findings (Section 3 and 5), we propose a framework (see 7) balancing artists' creative processes and viewers' diverse expectations to guide the design and development of art chatbots.

6.2.1 Artists and Chatbot Collaboration. Artists traditionally take the responsibility of answering viewer questions and fostering meaningful engagement. Based on the findings in Study 1, chatbots should complement an artist's role by handling factual or repetitive inquiries without impinging on creative processes or intellectual property. Several participants suggested periodically interviewing artists to **update the knowledge base** of a chatbot, ensuring the chatbot accurately represents each artist's evolving perspectives and remains true to their voice. Such updates can also help incorporate artists' narrative guidelines, so the chatbot knows when to flag more subjective or emotional questions. Moreover, most participants (artists) in Study 1 emphasized that interpretive or emotional questions demand a human touch; accordingly, our framework should include a dedicated **Artist Direct** channel that routes these subjective inquiries immediately to the artist for an authentic response. This collaborative balance preserves artistic integrity and supports a richer viewer experience.

This approach addresses the concern that chatbots might misinterpret deeply personal narratives—particularly those dealing with emotional backstories (See Section 3.2.2). By having regular updates with artists, galleries can provide chatbots with relevant context (e.g., new exhibitions, personal interpretations, style changes), allowing the chatbot to stay current without creating additional burdens on the artist. Ultimately, **interpretive inquiries** are routed back to the artist for a more authentic, human touch. This collaborative balance helps preserve artistic integrity and supports a richer viewer experience.

To protect the authenticity of each artist's voice, the system should offer an **Artist-Preview** Mode in which artists can audit chatbot answers that quote or interpret their statements before they are made public. This feature directly addresses artists' expressed worry that "personal context may be misinterpreted by a chatbot" (A2, A5 in Study 1) and complements the broader intellectual-property safeguards discussed in Section 6.2.4. Artists in our interviews (A1, A3) also welcomed the idea of receiving periodic updates that summarize the most frequent viewer questions. Such analytics can inspire future works or educational content while helping artists detect emerging misconceptions. Coupling this human-in-the-loop feedback mechanism ensures that the collection and reuse of aggregated viewer-question data respect intellectual-property rights while still empowering reflective practice.

6.2.2 Viewers and Chatbot Interaction. Based on the results from Study 2, most viewers directed factual and low-complexity questions to chatbots, while complex and interpretive inquiries remained the realm of artists. Consequently, chatbots should be designed to distinguish question types and complexity to help route viewers'

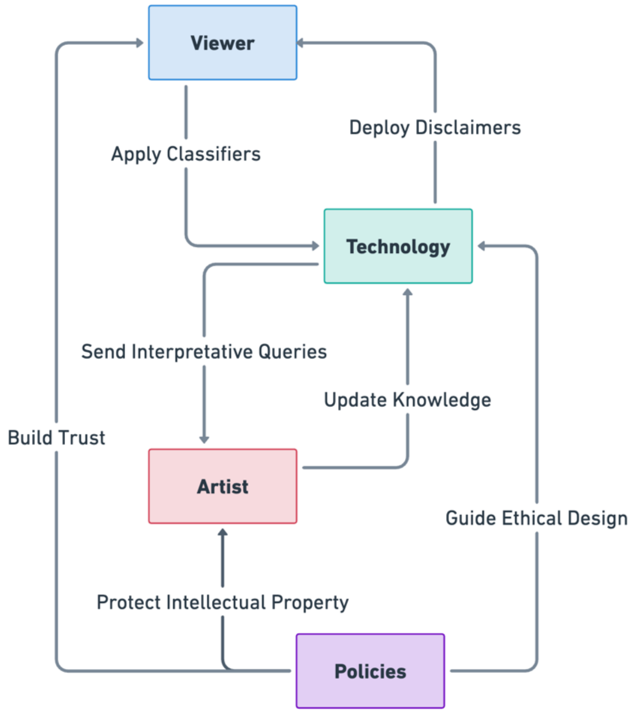

Figure 7: Framework for art chatbot design balancing different stakeholders.

inquiries efficiently. At the same time, some viewers expressed interest in more "human-like" chatbot conversations, underscoring the need for **disclaimers** that clarify the chatbot's role and limitations for those seeking deeper or more emotional discussions [49]. By dynamically adapting to different questions—through question-type classification (see Section 5.1) and complexity assessment (see Section 5.2)—chatbots can either provide partial information or recommend artist input for more nuanced matters. This adaptive approach reflects the diverse questioning tendencies observed in Section 5, offering both efficiency for routine inquiries and opportunities for richer engagement where human expertise is preferred.

Implementing the art chatbot can begin with a pilot launch of a minimal viable version that primarily handles factual queries. This approach allows for iterative improvements, informed by user feedback and error monitoring. For example, data on the frequencies of user question types can be tracked to identify which queries are consistently or frequently misrouted. Viewer satisfaction and response accuracy can also be monitored to guide refinements in both chatbot answers and question-routing logic.

6.2.3 Technology and Chatbot Functionality. Based on the findings from Section 3 on preserving authenticity, galleries should develop a curated, periodically updated repository of reliable data (e.g., official statements, artist-approved narratives), ensuring that chatbots accurately reflect artists' evolving views. This approach helps mitigate the "hallucinations" produced by LLMs and keeps factual data current. Informed by the analysis of question types and complexity in Section 5, a **classifier** trained on real viewer questions can help route queries to the appropriate respondent. A hybrid architecture may also be effective for balancing viewer expectations. The system can combine a rule-based layer for basic factual questions with an LLM-based layer for broader, open-ended queries. This dual approach upholds efficiency for straightforward queries, while accommodating more exploratory or nuanced topics that require richer engagement.

To further address the concern expressed by artists in Study 1 that personal narratives could be misconstrued, every chatbot answer should be given a score. When the score falls below a predefined threshold, the system either attaches a disclaimer or diverts the question to the Artist Direct channel described in Section 6.2.1. This mechanism also responds to viewers' preferences identified in Study 2 for clearly knowing when "human expertise" is involved. Aggregated logs should not only support error correction but also provide artists with the most frequently asked viewer questions; our interviewees (A1,A3) said such insights could spark ideas for future exhibitions or artworks.

Additionally, implementing real-time data logs and feedback loops enables galleries to monitor chatbot outputs, ensuring misinterpretations or inaccurate answers are promptly flagged and corrected. It is also critical to provide explicit fallback options for users who need deeper, more personal engagement. Such transparency not only **builds trust** with viewers but also reassures artists that the system is meant to complement their expertise rather than replace the authenticity of direct human interaction.

6.2.4 Ethical and Cultural Considerations. Ethical considerations in chatbot design are essential, particularly concerning user privacy and transparency. Findings from Section 3 suggest that the artists emphasized the need for control over how their creative processes and narratives are represented, aligning with broader AI ethical guidelines. Clear disclosure of the chatbot's AI nature helps set accurate viewer expectations [50], consistent with the observation in Section 5 that many viewers rely on artists for subjective responses. Additionally, safeguarding mechanisms such as moderation filters can prevent misuse, ensuring appropriate handling of sensitive topics by trained personnel.

Cultural sensitivity is also particularly salient in context-dependent domains like traditional Chinese painting, where motifs such as plum blossom and crane carry centuries-old symbolic weight. An LLM-based art chatbot trained predominantly on Western art corpora may overlook or misinterpret such connotations, reinforcing a 'Western-centric' reading. We therefore recommend incorporating curated, language-specific corpora or adding a culture-specific validation layer when accuracy in symbolic interpretation is important.

The art chatbot should present interpretive plurality by default. For example, when a viewer asks about the meaning of a crane motif, the chatbot provides both the classical interpretation and personal views gathered from artists. Because art interpretation is inherently subjective, each answer should state that it is one of multiple valid interpretations and provide an option to request the artist's personal view via Artist Direct. For derivative content, the chatbot should embed gallery-specified licenses in text unless the artist has granted explicit rights.

Beyond content moderation, data governance also requires attention. Galleries should specify how user data or questions might be stored, analyzed, and potentially shared, giving viewers a clear way to opt out if they wish. This transparency not only complies with privacy standards but also fosters user trust—an aspect particularly important when dealing with personal inquiries. These measures address concerns highlighted by our two studies, emphasizing the importance of protecting **intellectual property** and ensuring user privacy [51, 52].

6.3 Limitations and Future Work

Our study has certain limitations that point toward areas of future inquiry. Although our quantitative approach to collecting viewer questions was effective for identifying broad trends, it lacked deeper qualitative insights into viewers' personal motivations and experiences. A dedicated follow-up interview study with viewers could offer insights into their perceptions of interactions with chatbots in more detail. Trust in art chatbots also emerged as a significant concern, as evidenced by participants in Study 2 who reserved subjective or emotionally nuanced questions for artists, citing concerns about chatbots' ability to provide meaningful responses. This suggests that viewers may be less inclined to trust chatbots with interpretive or personal topics. Addressing this trust gap requires further research into strategies to improve chatbot transparency and contextual understanding. Moreover, because participants were aware that their questions would not be answered in real-time, this may have influenced their questioning behavior due to demand characteristics. Specifically, the absence of time pressure, since no respondent was waiting, may have led them to formulate their questions more deliberately or with greater elaboration. Nonetheless, future studies should compare question patterns collected in natural gallery tours to evaluate ecological validity.

Although focusing on traditional Chinese painting unveiled culture-specific challenges, our framework is intended to be extensible. Future studies should conduct cross-genre replications—e.g., installation or digital art—to test whether our question-routing logic and complexity classifier still hold when works are multi-modal or abstract. Comparative experiments could keep the chatbot architecture constant while varying the artwork corpus to measure shifts in question distribution and perceived trust. Cross-cultural adaptation also warrants systematic study. A promising design is to experiment with different art genres (e.g., European oil painting) with visitors from various cultural backgrounds. Dependent variables would include not only question type/complexity, but also cultural appropriateness described in Section 6.2.4. Such work can reveal whether additional culture-specific validation layers are needed.

The study's experimental design involved collecting questions without providing real-time chatbot or artist responses. While this approach simplified comparisons and isolated questioning behaviors, it prevented us from observing how dynamic chatbot feedback might shape follow-up questions. Building a functional prototype chatbot integrating the insights of this study and testing it in real or simulated gallery conditions will be a critical next step. Such empirical work could uncover how response quality, personalization, or user interface design influences viewers' continued engagement. We also did not systematically evaluate which specific LLM architectures are best suited for art-related Q&A tasks. Future work could measure multiple LLMs on their ability to answer art questions accurately, as well as assess their susceptibility to hallucinations when dealing with subjective or culturally nuanced content.

7 CONCLUSION

Art institutions worldwide are increasingly turning to LLMs and conversational agents to enrich viewer experiences. This paper contributes an initial and significant step toward realizing an art chatbot by exploring both artists' and viewers' perspectives on its potential roles and functionalities. Our findings highlight a distinct divide in questioning patterns: the majority of viewers regard chatbots as a source of factual or less complex information, while seeking emotional and interpretive answers directly from artists. Notably, a small group of viewers sees chatbots as capable of addressing more subjective issues, which underscores the need for adaptable chatbot designs.

By integrating insights from the artists with findings from the viewers, we developed a framework that ensures factual queries are efficiently handled by chatbots, while subjective or complex inquiries can be handed over to artists. Both artists and viewers largely support this distribution of labor, but neither side advocates a rigid division where certain types of questions bypass artists or chatbots entirely. The proposed framework (Artists, Viewers, Technology, Policies) provides a roadmap for balancing the needs of artists, viewers, and galleries in chatbot design. Its validation in real-world settings remains a key area for future work. Pilot implementations in galleries, with feedback from both artists and viewers, would strengthen the framework's practical applicability.

In summary, our study underscores the importance of both stakeholder alignment and system design when introducing chatbots into art contexts. This research sets the stage for further exploration of real-time chatbot-human interactions in diverse cultural settings and provides a stepping stone toward a new era of art engagement.

References

[1] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. 2024. A Comprehensive Overview of Large Language Models. https://doi.org/10.48550/arXiv.2307.06435

[2] Ujjwal Adhikari. 2023. Exploring the Painting Artist Perceptions Towards Their Career Sustainability in Pokhara. *J. Interdis. Stud.* 12, 1 (December 2023), 79–92. https://doi.org/10.3126/jis.v12i1.65450

[3] Michael Gerlich. 2024. Exploring Motivators for Trust in the Dichotomy of Human—AI Trust Dynamics. Social Sciences 13, 5 (May 2024), 251. https://doi.org/10.3390/socsci13050251

[4] Inas Alkholy. 2009. Is the Reception of Emotional Expression in Visual Art Global? The International Journal of the Humanities: Annual Review 7, 3 (2009), 179–188. https://doi.org/10.18848/1447-9508/CGP/v07i03/42651

[5] Manal Almalki. 2021. Exploring the Influential Factors of Consumers' Willingness Toward Using COVID-19 Related Chatbots: An Empirical Study. Med Arch 75, 1 (2021), 50. https://doi.org/10.5455/medarh.2021.75.50-55

[6] Taeha Yi, Mi Chang, Sukjoo Hong, Meereh Kim, and Ji-Hyun Lee. 2020. A Study on Understanding of Visitor Needs in Art Museum: Based on Analysis of Visual Perception Through Eye-Tracking. In Intelligent Human Systems Integration 2020, 2020. Springer International Publishing, Cham, 1132–1137. https://doi.org/10.1007/978-3-030-39512-4_172

[7] Sarah Cascone ShareShare This Article. 2018. How Hard Is It to Make It as an Artist? A New Study Shows There Are Many Paths to Success—and Failure. Artnet News. Retrieved December 28, 2024 from https://news.artnet.com/art-world/artist-financial-stability-survey-1300895

[8] Claire Bailey-Ross, Steven Gray, Jack Ashby, Melissa Terras, Andrew Hudson-Smith, and Claire Warwick. 2016. Engaging the museum space: Mobilizing visitor engagement with digital content creation. Digital Scholarship Humanities (October 2016), fqw041. https://doi.org/10.1093/llc/fqw041

[9] Austin Beattie, Autumn P. Edwards, and Chad Edwards. 2020. A Bot and a Smile: Interpersonal Impressions of Chatbots and Humans Using Emoji in Computer-mediated Communication. Communication Studies 71, 3 (May 2020), 409–427. https://doi.org/10.1080/10510974.2020.1725082

[10] Authors: Department of Asian Art. Landscape Painting in Chinese Art | Essay | The Metropolitan Museum of Art | Heilbrunn Timeline of Art History. The Met's Heilbrunn Timeline of Art History. Retrieved January 19, 2025 from https://www.metmuseum.org/TOAH/HD/clpg/hd_clpg.htm?utm_source$=$chatgpt.com

[11] Tamal T. Biswas and Kenneth W. Regan. 2015. Quantifying Depth and Complexity of Thinking and Knowledge: In Proceedings of the International Conference on Agents and Artificial Intelligence, 2015. SCITEPRESS - Science and and Technology Publications, Lisbon, Portugal, 602–607. https://doi.org/10.5220/0005288306020607

[12] Amy Hinsley, William J. Sutherland, and Alison Johnston. 2017. Men ask more questions than women at a scientific conference. PLoS ONE 12, 10 (October 2017), e0185534. https://doi.org/10.1371/journal.pone.0185534

[13] Lin Cheng, Junping Xu, and Younghwan Pan. 2023. Evaluating Influencing Factors of Visitor Experience in Offline Art Exhibitions. https://doi.org/10.20944/preprints202310.0845.v1

[14] Hyojin Chin, Gabriel Lima, Mingi Shin, Assem Zhunis, Chiyoung Cha, Junghoi Choi, and Meeyoung Cha. 2023. User-Chatbot Conversations During the COVID-19 Pandemic: Study Based on Topic Modeling and Sentiment Analysis. J Med Internet Res 25, (January 2023), e40922. https://doi.org/10.2196/40922

[15] Hyojin Chin, Hyeonho Song, Gumhee Baek, Mingi Shin, Chani Jung, Meeyoung Cha, Junghoi Choi, and Chiyoung Cha. 2023. The Potential of Chatbots for Emotional Support and Promoting Mental Well-Being in Different Cultures: Mixed Methods Study. J Med Internet Res 25, (October 2023), e51712. https://doi.org/10.2196/51712

[16] Sharing the Depth of Knowledge Framework With Students. Edutopia. Retrieved January 6, 2025 from https://www.edutopia.org/blog/sharing-depth-knowledge-wheel-students-johanna-tramantano

[17] Ahmed Elgammal, Marian Mazzone, Bingchen Liu, Diana Kim, and Mohamed Elhoseiny. 2018. The Shape of Art History in the Eyes of the Machine. https://doi.org/10.48550/ARXIV.1801.07729

[18] Mohammad Amin Kuhail, Mohamed Bahja, Ons Al-Shamaileh, Justin Thomas, Amina Alkazemi, and Joao Negreiros. 2024. Assessing the Impact of Chatbot-Human Personality Congruence on User Behavior: A Chatbot-Based Advising System Case. IEEE Access 12, (2024), 71761–71782. https://doi.org/10.1109/ACCESS.2024.3402977

[19] Duplicate: Stavros Vassos, Eirini Malliaraki, Federica dal Falco, Jessica Di Maggio, Manlio Massimetti, Maria Giulia Nocentini, and Angela Testa. 2016. Art-Bots: Toward Chat-Based Conversational Experiences in Museums. In Interactive Storytelling, 2016. Springer International Publishing, Cham, 433–437. https://doi.org/10.1007/978-3-319-48279-8_43

[20] Ben F. Cotterill. 2023. The Big Five. In Personality Psychology, Ideology, and Voting Behavior: Beyond the Ballot, Ben F. Cotterill (ed.). Springer Nature Switzerland, Cham, 9–28. https://doi.org/10.1007/978-3-031-39642-7_2

[21] Niklas Humble and Peter Mozelius. 2022. Content analysis or thematic analysis: Similarities, differences and applications in qualitative research. ecrm 21, 1 (May 2022), 76–81. https://doi.org/10.34190/ecrm.21.1.316

[22] Todd B. Kashdan, Matthew W. Gallagher, Paul J. Silvia, Beate P. Winterstein, William E. Breen, Daniel Terhar, and Michael F. Steger. 2009. The curiosity and exploration inventory-II: Development, factor structure, and psychometrics. Journal of Research in Personality 43, 6 (December 2009), 987–998. https://doi.org/10.1016/j.jrp.2009.04.011

[23] Yoonsu Kim, Jueon Lee, Seoyoung Kim, Jaehyuk Park, and Juho Kim. 2023. Understanding Users' Dissatisfaction with ChatGPT Responses: Types, Resolving Tactics, and the Effect of Knowledge Level. (2023). https://doi.org/10.48550/ARXIV.2311.07434

[24] Ana Paula Chaves and Marco Aurelio Gerosa. 2021. How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. International Journal of Human–Computer Interaction 37, 8 (May 2021), 729–758. https://doi.org/10.1080/10447318.2020.1841438

[25] Art Business News. 2020. How is Chatbot Technology Changing the Art Scene? Art Business News. Retrieved December 28, 2024 from https://dev.artbusinessnews.com/2020/11/how-is-chatbot-technology-changing-the-art-scene/

[26] Abiodun Finbarrs Oketunji, Muhammad Anas, and Deepthi Saina. 2023. Large Language Model (LLM) Bias Index—LLMBI. https://doi.org/10.5281/ZENODO.10441700

[27] How to Engage Gifted Students in the Classroom | Theory & Tips. Retrieved January 6, 2025 from https://slideswith.com/blog/how-to-engage-gifted-students

[28] Lois H. Silverman. 1995. Visitor Meaning-Making in Museums for a New Age. Curator The Museum Journal 38, 3 (September 1995), 161–170. https://doi.org/10.1111/j.2151-6952.1995.tb01052.x

[29] Jim Richardson. 2019. How Are Museums Using Chatbots? MuseumNext. Retrieved December 30, 2024 from https://www.museumnext.com/article/how-are-museums-are-using-chatbots/

[30] Jim Richardson. 2022. A Conversational AI Guide at Centre Pompidou. MuseumNext. Retrieved December 28, 2024 from https://www.museumnext.com/article/a-conversational-ai-guide-at-centre-pompidou/

[31] Robert Kozinets, Ulrike Gretzel, and Anja Dinhopl. 2017. Self in Art/Self As Art: Museum Selfies As Identity Work. Front. Psychol. 8, (May 2017), 731. https://doi.org/10.3389/fpsyg.2017.00731

[32] Hanna K. Schraffenberger and Edwin Van Der Heide. 2015. Audience-artwork interaction. IJART 8, 2 (2015), 91. https://doi.org/10.1504/IJART.2015.069550

[33] Siddhant Bikram Shah, Surendrabikram Thapa, Ashish Acharya, Kritesh Rauniyar, Sweta Poudel, Sandesh Jain, Anum Masood, and Usman Naseem. 2024. Navigating the Web of Disinformation and Misinformation: Large Language Models as Double-Edged Swords. IEEE Access (2024), 1–1. https://doi.org/10.1109/ACCESS.2024.3406644

[34] Cheung-On Tam. 2022. Dialogue-and-Questioning Strategies in Art Classrooms and Museums. Studies in Art Education 63, 3 (July 2022), 202–219. https://doi.org/10.1080/00393541.2022.2081029

[35] Lady Joanne Tjahyana. 2024. Exploring AI Chatbot Development for Gen-Z: A Study on First-Time and Experienced Voters in Pemilu. JIKB 9, 2 (April 2024), 224–236. https://doi.org/10.36914/jikb.v9i2.1091

[36] Xiaoming Zhai, Matthew Nyaaba, and Wenchao Ma. 2024. Can Generative AI and ChatGPT Outperform Humans on Cognitive-Demanding Problem-Solving Tasks in Science? Sci & Educ (January 2024). https://doi.org/10.1007/s11191-024-00496-1

[37] Ben Walmsley. 2016. From arts marketing to audience enrichment: How digital engagement can deepen and democratize artistic exchange with audiences. Poetics 58, (October 2016), 66–78. https://doi.org/10.1016/j.poetic.2016.07.001

[38] Liang Wenjing and Nadia Mohd Nasir. 2023. Principles and Practice of Landscape Design: An Application of Traditional Chinese Landscape Painting. GJESET 1, 1 (May 2023), 12–18. https://doi.org/10.56225/gjeset.v1i1.16

[39] Yingyi Hao, Mohd Fabian Hasna, and Faziawati Abdul Aziz. 2024. The role of emotional mediation in visitor experiences in art tourism: based on stimulus-organism-response model. Curr Psychol 43, 44 (November 2024), 34018–34034. https://doi.org/10.1007/s12144-024-06913-6

[40] Peixiang Zhong, Chen Zhang, Hao Wang, Yong Liu, and Chunyan Miao. 2020. Towards Persona-Based Empathetic Conversational Models. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020. Association for Computational Linguistics, Online, 6556–6566. https://doi.org/10.18653/v1/2020.emnlp-main.531

[41] Payal Arora and Filip Vermeylen. 2013. THE END OF THE ART CONNOISSEUR? EXPERTS AND KNOWLEDGE PRODUCTION IN THE VISUAL ARTS IN THE DIGITAL AGE. Information, Communication & Society 16, 2 (March 2013), 194–214. https://doi.org/10.1080/1369118X.2012.687392

[42] 2023. Research on the Aesthetic Value and Artistic Characteristics in Chinese Painting. artpl 4, 4 (2023). https://doi.org/10.23977/artpl.2023.040410

[43] 2023. Inheritance and Development—Thinking and Exploration on the Current Situation of Chinese Meticulous Figure Painting. artpl 4, 2 (2023). https://doi.org/10.23977/artpl.2023.040214

[44] Mike Murawski. 2015. Questioning the Questioning of Questions. Art Museum Teaching. Retrieved January 5, 2025 from https://artmuseumteaching.com/2015/01/23/questioning-the-questioning-of-questions/

[45] Akron Art Museum. Retrieved December 28, 2024 from https://www.theformgroup.com/work/akron-art-museum

[46] Send Me SFMOMA. SFMOMA. Retrieved December 28, 2024 from https://www.sfmoma.org/read/send-me-sfmoma/

[47] Tingting Liu, Salvatore Giorgi, Ankit Aich, Allison Lahnala, Brenda Curtis, Lyle Ungar, and João Sedoc. 2024. The Illusion of Empathy: How AI Chatbots Shape Conversation Perception. https://doi.org/10.48550/arXiv.2411.12877

[48] Akiko Walley. 2023. Artworks and Art History: Toward a Deeper Engagement with Art Exhibition and/as Art. In Visiting the Art Museum: A Journey Toward Participation, Eleonora Redaelli (ed.). Springer International Publishing, Cham, 87–108. https://doi.org/10.1007/978-3-031-12089-3_6

[49] Luise Metzger, Linda Miller, Martin Baumann, and Johannes Kraus. 2024. Empowering Calibrated (Dis-)Trust in Conversational Agents: A User Study on the Persuasive Power of Limitation Disclaimers vs. Authoritative Style. In Proceedings of the CHI Conference on Human Factors in Computing Systems, May 11, 2024. ACM, Honolulu HI USA, 1–19. https://doi.org/10.1145/3613904.3642122

[50] Jessica Y. Bo, Harsh Kumar, Michael Liut, and Ashton Anderson. 2024. Disclosures & Disclaimers: Investigating the Impact of Transparency Disclosures and Reliability Disclaimers on Learner-LLM Interactions. HCOMP 12, (October 2024), 23–32. https://doi.org/10.1609/hcomp.v12i1.31597

[51] Lawrence Lessig. 2005. Free culture: the nature and future of creativity (Reprint ed.). Penguin Books, New York, NY.

[52] Helen Nissenbaum. 2004. Privacy as Contextual Integrity. Washington Law Review 79, 1 (February 2004), 119. Retrieved from https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10

[53] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33, 1 (March 1977), 159. https://doi.org/10.2307/2529310

[54] Guendalina Caldarini, Sardar Jaf, and Kenneth McGarry. 2022. A Literature Survey of Recent Advances in Chatbots. Information 13, 1 (January 2022), 41. https://doi.org/10.3390/info13010041

[55] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. https://doi.org/10.48550/ARXIV.2005.14165

Curiosity and Exploration Inventory–II (CEI–II)

A Instructions

Please indicate your level of agreement with each statement using the following rating scale:

- 1 = Strongly Disagree
- 2 = Disagree
- 3 = Neither Agree nor Disagree
- 4 = Agree
- 5 = Strongly Agree

There are no right or wrong answers; please respond as honestly and accurately as possible.

A.1 Items

- I actively seek as much information as I can in new situations.
- I am the type of person who really enjoys the uncertainty of everyday life.
- I am at my best when doing something that is complex or challenging.
- I prefer jobs that are excitingly unpredictable.
- I frequently seek out opportunities to challenge myself and grow as a person.
- I am the kind of person who embraces unfamiliar people, events, and places.
- Everywhere I go, I am out looking for new things or experiences.
- I like to do things that are a little frightening.
- I am always looking for experiences that challenge how I think about myself and the world.
- I view challenging situations as an opportunity to grow and learn.

B Webb's Depth of Knowledge (DOK) Level and Criteria

C Additional Logistic Regression Results

DOK Level	Definition	Art Appreciation Focus
Level 1 Recall & Reproduction	Requires recall of facts or rote application of simple procedures. The task involves straightforward recognition or recall of basic information.	Simple identification and recall: Viewers recall specific, surface-level details about the artwork (e.g., title, artist's name, medium, date of creation).
Level 2 Skills & Concepts	Involves more than one mental step; typically requires conceptual understanding, classification, or comparing/contrasting.	Interpretation with basic comparisons: Viewers make observations or straightforward connections between elements of an artwork (e.g., comparing color palettes, noting compositional features).
Level 3 Strategic Thinking	Requires reasoning, planning, and using evidence. Questions at this level ask for deeper analysis, such as explaining motives or ideas.	In-depth analysis and explanation: Viewers explore how meaning is conveyed, how artists' context or technique influences interpretation, or how the artwork relates to broader themes.
Level 4 Extended Thinking	Demands extended planning, investigation, or critical synthesis over time; tasks often involve evaluating and reflecting on multiple sources.	Critical synthesis and personal evaluation: Viewers generate comprehensive critiques, integrate various perspectives (historical, cultural, emotional), and articulate a personalized stance that goes beyond straightforward analysis.

Table 7: Multinomial Logistic Regression Results of IP Category (Reference Category: AF)

Predictor	Coefficient	p-values	OR (95% CI)
Intercept	0.402	0.685	1.495 [0.207, 10.823]
Curiosity Score	-0.036	0.241	0.965 [0.908, 1.025]
Age	-0.006	0.276	0.994 [0.983, 1.005]
Expected Respondent=Artist	**1.046**	**<0.001**	**2.845 [1.947, 4.159]**
Gender=Female	-0.124	0.534	0.883 [0.597, 1.307]
Education=Postgraduate	0.170	0.576	1.185 [0.654, 2.145]
Education=Pre-University	0.266	0.225	1.304 [0.849, 2.004]
Gallery Visiting Frequency=Once a month or more	-0.414	0.130	0.661 [0.387, 1.129]
Gallery Visiting Frequency=Once a year or less	-0.173	0.483	0.841 [0.519, 1.363]
Gallery Visiting Frequency=Once around half a year	-0.423	0.099	0.655 [0.396, 1.082]
Painting Type = Bird and flower painting	-0.107	0.638	0.899 [0.575, 1.404]
Painting Type = Figure painting	0.304	0.174	1.356 [0.874, 2.103]

The reference categories: Expected Respondent = Chatbot, Gender = Male, Education = University, Gallery Visiting Frequency = Once around three months, Painting Type = Landscape painting

Table 8: Multinomial Logistic Regression Results of AE Category (Reference Category: AF)

Predictor	Coefficient	p-values	OR (95% CI)
Intercept	−1.422	0.306	0.241 [0.024, 2.417]
Curiosity Score	−0.018	0.668	0.982 [0.903, 1.068]
Age	0.002	0.779	1.002 [0.988, 1.017]
Expected Respondent=Artist	**0.395**	**0.132**	**1.484 [0.887, 2.483]**
Gender=Female	0.167	0.548	1.182 [0.685, 2.040]
Education=Postgraduate	0.265	0.517	1.303 [0.585, 2.902]
Education=Pre-University	0.373	0.225	1.452 [0.795, 2.653]
Gallery Visiting Frequency=Once a month or more	-0.346	0.381	0.707 [0.326, 1.534]
Gallery Visiting Frequency=Once a year or less	-0.135	0.692	1.144 [0.588, 2.225]
Gallery Visiting Frequency=Once around half a year	-0.064	0.854	0.938 [0.473, 1.859]
Painting Theme = Bird and flower painting	-0.080	0.795	0.923 [0.505, 1.686]
Painting Theme = Figure painting	0.089	0.775	1.093 [0.595, 2.005]

The reference categories: Expected Respondent = Chatbot, Gender = Male, Education = University, Gallery Visiting Frequency = Once around three months, Painting Type = Landscape painting