Current Topics and Developments in HCI II
Study Program - Usability Engineering (M.Sc.)
Semester - WS 2024/25
Documentation & Manuscript

Museums Reimagined: A Multimodal Chatbot Prototype to Foster Young Visitors' Engagement with Contemporary Art

Student Name - Matriculation Number:
Itzel G. Vizcaíno García - 35481
Kübra Aslan - 36237
Ümmü Gülsüm Ergin - 36270

[Citation standard: APA-7]

Table Of Contents

Table of Figures

---

1. Introduction

1.1 Background

Within the changing shift towards digitalization, museums have been integrating emerging technologies such as gamification, social media, non-fungible tokens (NFTs), augmented reality (AR) applications, virtual reality (VR), and mixed reality (MR) into their exhibitions to attract more visitors and enhance their visitor experience (Giannini & Bowen, 2022). Integrating emerging technologies remains an ongoing process, accompanied by a range of advantages and disadvantages for each approach. Within the research concluded by the team members' previous semester, chatbots' potential to optimize the visitor experience and create interactive experiences while also keeping implementation economically efficient and technologically feasible is regarded as a valuable opportunity. This perspective is further supported by the findings of Bordoni et al. (2016) and Chai-Arayalert et al. (2024) on the utilisation of chatbots in contrast to alternative models. In the literature and existing museum practices, as given in the several examples above, a diversity of types and scopes of museums is covered. Nevertheless, there is an existing gap in the implications of chatbots for contemporary art exhibitions. Subsequently, research has shown that young people have a low threshold for visiting museums (Cantini, 2015; Drotner et al.,2017; Pierroux et al.,2011). The underlying reasons for the lack of interest among youth are boredom associated with museums and heavy information, with a lack of interaction (Drotner et al.,2017). Given the current body of information, the need for further studies that concentrate on young adults has been identified. In addition, most of these studies were conducted within a 10-year period and have limited capacity to reflect current conditions, changing youth characteristics, and evolving technological improvements. The characteristics and expectations of young groups have changed significantly in the post-pandemic period (Giannini & Bowen, 2022). Therefore, in this study, individuals between the ages of 15-24 as of 2024 were determined as the target audience, based on the United Nations' age classification. The development of a multimodal AI chatbot has been initiated by this foundation, and collected knowledge has been collected from prior literature research conducted.

1.2. Research Objectives

This study aims to construct a secure, multimodal AI-powered conversational agent (chatbot) using an interactive prototype to support a museum experience, to enhance the engagement level by targeting young visitors and their needs. Moreover, another aim of the study is to test the validity of the developed solution by conducting testing in a real museum setting. Specifically for chatbot design, it has been prioritized to create an intuitive, user-centered interface design enriched by engaging interactions and supporting diversified formats such as voice, text, and multimedia. With the help of a developed prototype design, the ultimate aim of the study is to have an increased level of usability and enhanced visitor engagement.

1.3. Scope and Research Questions

The following question is intended to be responded to within the scope of this project: How can AI chatbots in a contemporary art exhibition increase young visitors' understanding of the stories behind artifacts?

Furthermore, the primary objective of the project is the development of a secure, multimodal AI-powered chatbot for museums, with the aim of enhancing engagement levels. The target demographic for the proposed solution is defined as young visitors. In order to achieve this objective, it is necessary to consider the needs and preferences of visitors to the museum who fall within the 15-24 age bracket. Furthermore, the user experience of the chatbot design is intended to be intuitive and seamless, with the principles of human-centered design being applied in order to achieve this objective. Finally, the validity of the solution is to be ascertained through simulated museum environment testing. This will entail the collection of user feedback and the measurement of the success of the chatbot in comparison to traditional methods of information reception, such as brochures, in terms of increasing engagement levels.

1.4 Hypothesis

Within the proposal of the study, in order to assert diversified aspects of the research question, a set of hypotheses has been generated. These hypotheses seek to answer the question of the potential AI chatbot utilization in museum settings to enhance the engagement level of young visitors. First and foremost, according to a simple hypothesis, it is expected that there will be a positive correlation between the

engagement level of the visitors and AI chatbot usage. Hypothesis suggests that young visitors who engage through an AI chatbot tend to show a higher level of understanding of the artifacts in the exhibition compared to those who do not and use the brochure only as a source of information. On the other hand, according to the null hypothesis (H_0) indicates that the impact of using AI chatbots during sessions would cause no significant difference, and both groups show the same level of engagement regardless of chatbot integration or brochure. Lastly, the alternative hypothesis (H_1) predicts that the group that benefited from the AI chatbot during their sessions grasps significantly more information regarding the artifacts and a considerably greater level of engagement in contrast with the group without any digital solution. Consequently, these three hypotheses must be tested within the scope of the study to be approved or disproved.

2. Literature Review

2.1 Introduction

The 2022 Extraordinary General Assembly of ICOM (International Council of Museums) defined museums as: “a not-for-profit, permanent institution in the service of society that researches, collects, conserves, interprets, and exhibits tangible and intangible heritage. Open to the public, accessible, and inclusive, museums foster diversity and sustainability.” This definition emphasizes museums’ roles as social institutions committed to societal service and cultural preservation and their primary mission is to educate, research, and improve knowledge dissemination within communities.

Due to their character of social institutions, museums face the challenge of understanding and adapting their tools to the younger generations, since these generations “will decide the fate of this kind of institution in the not-too-distant future” (Kisiel, 2021, P. 11). Hence, the ability to conduct their mission in the society of these institutions relies on their ability to reach a defined target audience and satisfy the expectations of young visitors (Kisiel, 2021).

2.2 Museum Experience Design

Nor and Razak (2021) suggested that museums are optimal environments for testing new interaction techniques that will enhance visitor experience positively to provide

effective learning outcomes and enrich visitor experience. In particular, Karayılanoğlu and Arabacıoğlu (2020) acknowledge this changing shift in museum experience design as a result of accessible information and interactive experiences derived from recent communication innovations, which shifted these spaces to not only observation spaces but also experience spaces and attribute them to new dimensions.

A great deal of research has been conducted on the enrichment of digital native visitors' experience in a museum context. The existing body of works offers interaction design solutions powered with gamification, social media, non-fungible tokens (NFTs), augmented reality (AR) applications, virtual reality (VR), and mixed reality (MR) to improve visitor experiences (Giannini & Bowen, 2022). Some of these implications mentioned in their research include the 2018 Modigliani exhibition at the Tate Modern, which benefits VR, exhibitions employing mixed reality to redesign Van Gogh and Claude Monet's paintings from the l'Atelier des Lumières in Paris, and The London Barbican Centre's 2019 exhibition More than Human… These examples mainly focused on spatial design and creating 3D visual solutions, not building strong community engagement and a user-centric approach. However, the integration of these technologies presents challenges. Chai-Arayalert et al. (2024) point out the disadvantages of working with VR and AR tools as their costly burden and the tangled technical aspects, along with the varying familiarity of visitors with those methods.

Giannini and Bowen (2022) point out chatbots as one possible opportunity to create interactive and gamified experiences as a low-cost and user-centred interaction tool; they present a chance to improve user engagement, particularly among young visitors. Vassos et al. (2016) concur with this approach by asserting that the usage of chatbots in art museums is so-called "Art-bots". When they conducted their research almost a decade ago, chatbots were just emerging, they noted that with this growing interest in chatbots, it developed into an alternative medium for communication. To provide a baseline approach, they developed a solution for the Mario Praz Museum in Rome, which allows visitors to interact with the stories about the writer/collector (Mario Praz) and pieces acquired from the late 18th century and the beginning of the 19th century. This early implication example showcases possible prospects of native museum applications. Similarly, research on the implication of chatbots on a bigger scope was conducted by Boiano et al. (2018) and they illustrated examples powered

by Facebook Messenger such as the Anne Frank House museum chatbot that allows the visitor to discover the story of her and the Museum of Australian Democracy in Canberra which targets children to learn more over a right of voting and its history. Furthermore, Chai-Arayalert et al. (2024) propose a chatbot design to improve visitor engagement in The Namuensri Textile Museum. They conducted the system usability scale (sus) questionnaire that assessed teenagers between 15-24 years who are skilled in using smartphones and familiar with chatbots to further measure their satisfaction with them. Hence, the average sus score is revealed as 74.93, which is more than 6 points above the average number of 68 (see Figure 1). This study reveals the positive perspective of young users towards possible chatbot applications.

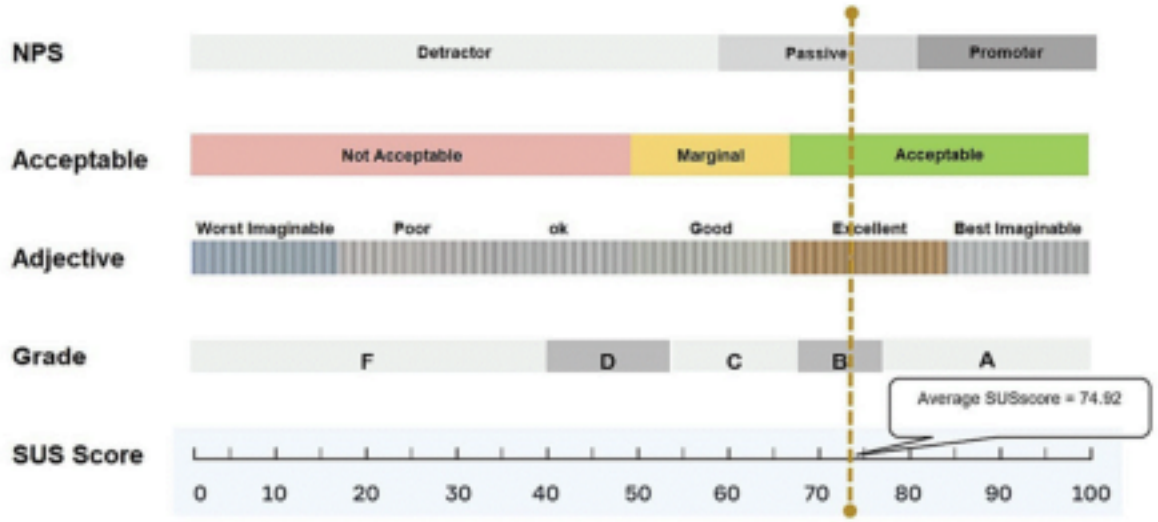


Figure 1: The results in the sus scale (Chai-Arayalert et al., 2024)

In the literature and existing museum practices, as given in the several examples above, a diversity of types and scopes of museums is covered. Nevertheless, there is an existing gap in the implications of chatbots for contemporary art exhibitions.

Indeed, Chai-Arayalert et al. (2024) stated that for robust chatbot design target audience profile and the unique context of museums should be considered. Therefore, to design a chatbot for a contemporary art exhibition, it is crucial to define the necessities of visitor groups and the distinctive characteristics of modern art exhibitions. Karayılanoğlu and Arabacıoğlu (2020) suggested that contemporary art museums could benefit from interaction design by enabling information transfer in an enjoyable, interactive, and informative manner. Moreover, they stressed the display of the artefacts as the primary focus of museology. Within this perspective, curators and

museum directors design exhibitions taking into account the concepts, artefacts, and targeted visitor groups. Miller (2002) clarifies the factors that have an influence over the meaning-making process from artefacts as local meanings, cultural appropriation, tradition, valuation, and cultural dynamics, along with the context and how they are displayed to the public. Hence, benefiting from interaction design for displaying artefacts could enable exhibitions to communicate with visitors effectively. Tsita et al. (2023) conducted research on this regard by working integration of augmented and virtual reality (VR) solutions to increase learning in contemporary art museums. They proposed VR interaction implications for the Metropolitan Organisation of Museums of Visual Arts of Thessaloniki (MOMus–Museum of Contemporary Art), where they curated the museum artefacts with four types of interactive activities, such as virtual guidance, 2D/3D animation, detailed insights, and hands-on activity. To validate their study, they tested their prototype with 3 different user groups, which can be named museum experts, technical experts, and general audiences. In the first stage of the assessment overall application was asserted by the museum experts who work in the MOMus exhibition. Secondly, technical feasibility testing is conducted by the experts to detect bugs and enhance user experience. Finally, a revised version was tested with the target audience, who were defined as people above 13 years old as end users of the application. Thus, the findings showed a positive trend in the desirability, feasibility, usability, and visitors' interest in contemporary art, acknowledged as improved. They revealed the increased level of visitors by stating, “The results were positive, with 87.5% positive responses (GA—Q4.1: 62.5% strongly agree; 25% agree).” However, as a limitation of the study, the testing of the learning effect of the proposed solution was pointed out as outside of the scope of the study, since a modified evaluation methodology is required. Thus, even though the existing body of work on the contemporary exhibitions context exists, there is a need to focus on the usage of chatbots in contemporary exhibitions and their learning impact.

Subsequently, research has shown that young people have a low threshold for visiting museums (Cantini, 2015; Drotner et al.,2017; Pierroux et al.,2011). The underlying reasons for the lack of interest among youth are boredom associated with museums and heavy information, with a lack of interaction (Drotner et al.,2017). Given the current body of information, the need for further studies that concentrate on young adults has been identified. In addition, most of these studies were conducted within a

10-year period and have limited capacity to reflect current conditions, changing youth characteristics, and evolving technological improvements. The characteristics and expectations of young groups have changed significantly in the post-pandemic period (Giannini & Bowen, 2022). Therefore, in this study, individuals between the ages of 15-24 as of 2024 were determined as the target audience, based on the United Nations' age classification.

To sum up, the existing body of work on HCI and museum context already covers implications of diversified solutions such as gamification, social media, non-fungible tokens (NFTs), augmented reality (AR) applications, virtual reality (VR), and mixed reality (MR) to improve visitor experiences (Giannini & Bowen, 2022).In the existing body of literature, researchers examined the implications of these interaction methods however, chatbots remained underappreciated. Additionally, chatbots are suggested as an economical and feasible solution for museums, and existing implications are examined by Vassos et al. (2016) and Boiano et al. (2018). Moreover, Chai-Arayalert et al. (2024) proposed a chatbot solution for a textile museum and gathered positive elicit from visitors aged between 15-24. Similarly, studies on the utilization of chatbots in contemporary art exhibitions are under-covered. Moreover, demographically young adults are named as the visitor group that has the lowest visitation frequency (Drotner et al.,2017). Consequently, the research endeavours to understand the utilization of AI chatbots in a contemporary art exhibition context to elevate young visitors' understanding of the stories related to artefacts are identified.

2.4 Technical Aspect of Museum Experience Design

Museums have used chatbots in order to assist visitors during the last decade (Varitimiadis et al., 2020). Adamopoulou and Moussiades (2020) investigated the central role of such chatbots is to respond simple visitor queries by either keyword matching. Another way to do this is by using the Knowledge Graphs (KG) approach (Varitimiadis et al., 2021). These chatbots are typically based on cutting-edge AI technologies including machine learning (ML), natural language processing (NLP), and the semantic web (SW) Varitimiadis et al., 2021).

The chatbot should be simple and easy to use, this is important because visitors should have easy access to the museum's content and not be confused with complicated technologies (Varitimiadis et al., 2020). Over time, chatbot usage in

museums changed (Tsitseklis et al., 2023). One early example is the work done with Max, a conversational agent that was installed in the HNF Museum in Germany in 2004. In this museum, visitors could use a keyboard to ask questions, and predetermined rules would determine the answers (Kopp et al., 2005). Chatbots which are being developed for navigating through cities that have a rich cultural background, for example, the one being developed for Naples, an encoder-decoder scheme applies for generating the answers (Sperlí, 2020).

A chatbot for online navigation in the museum was developed for the Museum of Palaeontology and Geology at Athens, helping a visitor understand various exhibits with multimedia information. This also recommends various related exhibits using a hybrid recommender system that takes inputs from both content-based and collaborative filtering. There are many NLP tasks employed in the system. These include but are not limited to name entity recognition in both English and Greek. All the functions are aided by a Knowledge Graph database (Tsitseklis et al., 2023).

Varitimiadis et al. (2020) highlight that, to have an effective chatbot in a museum environment, it needs to combine the features of NLP with continuous learning by ML. All these must be supported by Text Recognition and Speech Recognition to be able to react to voice input and text messages. Moreover, the system must be user-friendly, graphically appealing, while at the same time technically robust so as not to create complications for the visitor.

The Natural Language Understanding (NLU) kernel of the system is implemented with RASA NLU, which provides a framework for developing a new project that implements the functionalities of Rasa Technologies (Duguleană et al., 2020). A comparison was made between Rasa NLU and Neural Networks, and it has been foundthat Rasa NLU with a Conditional Random Fields (CRF) classifier did better in comparison with the Neural Network approach with a single hidden layer and an extra representation layer (Tsitseklis et al., 2023). In accordance with the subjects of interest of museum visitors, the RASA NLU core was designed to handle more than 300 queries organized into three primary categories: inquiries concerning the city, the museum as a whole, and its exhibits. One open question was left at the end of the research questionnaire. Of the more than 250 respondents, only 45 answered 25%. From the analysis of the research results relating to the 250 students who benefited from the visit of the temporary exhibition and who interacted with IA (Virtual

Assistant), one can deduce that, in general, the system was well accepted by the target audience (Duguleană et al., 2020).

Wang (2024) stated that another study investigated the differences between the use of a chatbot tour guide and a traditional museum application in an art museum, in relation to how a chatbot enhances the experience of visitors effectively. In this study, a total of 30 participants were considered: 14 females, 15 males, and 1 individual who identified as non-binary, between 17 and 64 years of age (M = 30.0, SD = 13.7). The participants were split into two groups, with one testing the chatbot and the other using the application. Seventeen participants reported that they visit museums more than five times yearly while four visit three to four times. On the other hand, nine visited less than twice a year. Moreover, 20 participants felt that they are professional-level art appreciators, and five of them like going around museums by themselves without a guided tour.

Effectiveness of the chatbot and museum app were measured using several measures. These included total visit duration, number of audio interactions, and location-based queries by visitors. Other measures within the chatbot interactions included photo-taking, message types of usage, and further questions asked. Content presentation was rated in terms of text and audio length, voice tone, and speed; the chatbot experience was rated regarding the naturalness of the conversation and preferences for different types of messages. Finally, participants had to rate their overall tour experience in terms of ease of finding information, learning about the artworks, interactivity, and enjoyment on a 5-point Likert scale (Wang, 2024).

In addition, Wang (2024) reported that participants spent an average of 25 minutes using the application and 23 minutes with the chatbot, viewing an average of 6.8 artworks on the app and 9.1 artworks with the chatbot. The chatbot has a high completion rate of 77%, while that for the app is only 50% realized. The mean number of photos clicked by the chatbot users is 7, with up to eight interactions with the chatbot at different frequencies: message, audio, and both. Participants reported that with the chatbot, finding information was easy, and experiences in both the app and chatbot were well enjoyed, though not significantly different in both. The length of the text/audio of both the app and the chatbot was found suitable; the voice tone was pleasant, and the audio speed was appropriate, likewise, all without significant differences with both conditions. The developers should be aware of the chatbot's

abilities in conversational AI and the limitation of these abilities, considering what they want the chatbot to deliver (Varitimiadis et al., 2020). Semantic Web technologies allow creating data and information repositories on the Web, developing common shared vocabularies, and providing formal rules for managing linked data (Cahn, 2017). SW technologies provide the much-needed structured information and knowledge representation that the chatbot technology needs to facilitate proper, machine-readable, machine-comprehensible dialogues, enabling them to answer human-like questions and responses as humans (Varitimiadis et al. 2020).

Chatbots should be traceable by the mobile application using GPS or any other way (Wang, 2024). By knowing the location of the visitor inside the museum, the chatbot will be able to offer him multimedia information according to his location inside the museum (Varitimiadis et al., 2020). Another issue that is very significant when dealing with chatbots is privacy and security since these applications are dealing with users' data and interactions (Varitimiadis et al., 2020).

These papers explore ways of enhancing visitor engagement with AI-driven chatbots in museum scenarios through NLP, ML, and Knowledge Graphs-to offer interactive, location-based, and personalized experiences. Such systems should be friendly, efficient, and technically sound, with respect for the privacy and security concerns, including the limits of AI. Given that current studies have already proven that generally speaking, chatbots do help visitors to become more engaged with the artifacts in an art museum, little is known about how such technologies facilitate the story-understanding task among the young audience of a modern art exhibit.

2.4 Psychological Aspect of Museum Experience Design

Art plays a central role in human societies by embodying culturally significant meanings. Its creation, representation, and appreciation occur within specific contexts that reinforce shared values or highlight the uniqueness of certain occasions. In the Western world, where art has been extensively institutionalized, aesthetic experience holds considerable value, as evidenced by substantial investments in its exhibition, conservation, and dissemination. To sustain and preserve this form of expression, institutionalization has primarily taken shape through museums (Šveb Dragija & Jelinčić, 2022). The significance and impact of these institutions in contemporary society lie in their ability to foster cultural exchange, as well as in their fundamental

role in safeguarding art, history, and heritage Museums are constantly trying to attract new audiences by developing policies, programs, and activities aimed at generating inclusive spaces and attracting new visitors (ICOM, 2017). Nevertheless, one group consistently reported as missing from museum visitor profiles in Western countries is young adults (Barron & Leask, 2017).

Within the context of their mission to disseminate knowledge, it is crucial for museums to comprehend and adapt their tools to younger generations. Kisiel (2021, p. 11) emphasizes that "younger generations participate in social and cultural life in a markedly different manner than previous generations." This gains particular importance when he states that "they will determine the future of such institutions in the near future." Some museums, particularly art museums, have adopted an educational approach enhanced with leisure-oriented elements to attract younger audiences. This includes integrating technology, such as mobile apps, to enrich the visitor experience to attract youth adults (Hughes & Moscardo,2019).

Mastandrea et al. (2007) conducted a study involving five hundred participants, in Italy, between 20 to 76 years old, to measure likeability and satisfaction with two distinct types of museum environments, comparing ancient and contemporary art collections. The study aimed to understand the motivations, expectations, and emotions driving museum visits. They found that 65.9% of participants visited contemporary art museums for "the pleasure of art," compared to 58.2% for ancient art museums. Additionally, the 2009 study by the same group of researchers aimed to provide a better understanding of the demographic and personality trait differences between visitors to museums of ancient art and those to museums of contemporary and modern art. In a sample of 137 Italian participants ranging from 19 to 69 years (83 participants from the Ancient Art Museum and 54 participants from the Modern and Contemporary Art Museum), the study revealed a difference in age, with visitors being about 10 years younger at GNAM compared to those at the Braschi Museum.

Even though the reasons behind visiting a museum may vary among visitors from the same generation, there is a social consensus that museums are informal learning institutions concerned with processes tied to education (Mastandrea et al., 2007, p. 174). The experience during the visit is crucial in motivating the learning of visitors. Liu and Idris (2018) propose segmenting the visiting process into three phases: Pre-visit, Visit, and Post-visit. This segmentation allows for a focused design of

services specific to each phase, tailored to the unique characteristics and objectives of the visitors' needs and motivations.

The different goals and motivations behind user behaviour shape their expectations and impact their level of satisfaction with the museum experience (Mastandrea et al., 2009). Understanding their goals, needs and generational characteristics can significantly improve user engagement. In this context Zheng and Liu (2024, p. 2) remark the importance of HCI in the enhancement of the visitor's experience: "By employing enjoyable and engaging HCI, digital museums of cultural heritage can better facilitate public access to cultural heritage information and enrich the overall user experience".

Recent advancements encompass a variety of human-computer interaction (HCI) instruments, including audio guides, chatbots and virtual reality to enhance the visitors' experience. The purposes of these technologies are varied. In this sense, Zheng and Liu (2024, p. 2) state that "effective HCI design in digital museums, through intuitive interaction and feedback mechanisms, significantly enhances user engagement with digital content." The creation and use of these tools underscore the importance of user experience (UX); Nielsen and Norman (1998) suggest that an exemplary user experience is when the exact needs of the customer are met, and in order to be a good experience, the products or services should exceed the expectations, bringing enjoyment. A successful UX includes how value is created for the visitor in their experiences.

Since most museums are considered educational environments rather than game, in many interventions' gamification concepts are introduced to enhance the UX, Liu and Idris (2018), although this introduce potential new challenges, as the risk of once the reward is removed, decreases, or changes its value for the user, there is a risk of behavioural extinction. Likewise, Nicholson (2012) says that designers should create systems that help users find their own reasons for engaging with the behaviour rather than rely on rewards.

Nicholson (2015) states that a way to encourage and maintain a behaviour is through the value perceived by the individual, where the Self-Determination Theory by Deci and Ryan (2004, as cited in Nicholson, 2015, p.3) identify three factors intrinsically linked to value and this is created in three phases: mastery—building confidence in

newfound knowledge or skills; autonomy—putting users in charge so they feel in control and can make their own choices; and relatedness—bringing people together who are doing these activities. Hence, the design of a valuable learning experience should enable users to master the knowledge in their own terms and relate with other users.

As Noh & Hong (2021, p.2) outline “Chatbots have been widely employed for users to understand user-centric interactions through situational awareness, personalization, and adaptation and can elicit a user’s self-disclosure through nonverbal dialogue behaviour, social dialogue, and social bond formation”. For instance, IA chatbots in museums can enable visitors' needs to have the choice to engage with the system and master the content on their own terms. In this aspect, Štekerová (2022) emphasizes the importance of understanding audience characteristics when designing chatbots, considering the language and generational expression patterns of the users.

In conclusion, if museums aim to preserve the historical and artistic heritage and continue with their mission of knowledge transfer while attracting younger generations, it is key to understand their audiences' characteristics and motivations for visiting exhibitions. Modern art museums represent two opportunity areas: their educational approach and the purpose of their audiences’ visit. Mastandrea et al. (2007) found that over 65% of visitors reported experiencing "the pleasure of art" at contemporary art museums. Additionally, Kisiel (2021) conducted a study with 396 Polish students between 16 to 20 years, to explore young audiences' expectations of contemporary art museums, reporting that 58.6% of respondents desired more information about museum objects, and 69.5% agreed that a video presentation next to artworks would make museums more interesting. This highlights the potential for Human-Computer Interaction (HCI) tools to enhance the understanding of artifacts.

Consequently, if the primary goal of a museum is to foster enduring behavioural change, chatbots can prioritize the enhancement of intrinsic motivation instead of offering instant rewards (Nicholson, 2015). This can be seen through the perceived cultural value of their interactions and its ability to connect users with their exhibitions (Zheng et al. 2024). Chatbots can be a tool to facilitate significant engagement over the medium to long term. When deciding the type of chatbot and its characteristics, there are different types of factors that should be considered. One of them is understanding the audience, which is important first to identify the intent of

the visitors and, second, to tailor the language used in creating conversational chatbots (Štekerová, 2022).

3. Chatbot Design and Content

In order to achieve the objective of enhancing the level of engagement exhibited by young visitors within the context of museums, the development of a chatbot was initiated. The primary function of this chatbot is to motivate visitors to explore alternative methods of information design, which deviate from conventional practices such as labels and brochures, by employing a conversational approach.

In the context of the chatbot, a total of four artworks from two artists have been selected to prioritise diversity with regard to artistic medium, cultural background, and subject matter. The following factors were considered during the selection process: the creation of diverse and distinctive voices for the chatbot and the establishment of a unique voice for the chatbot. Therefore, in order to ensure a diverse range of perspectives, two contemporary artists of different life experiences, cultural backgrounds, and historical contexts have been selected as follows: Frida Kahlo and Jean-Michel Basquiat.

Frida Kahlo can be named as one of the most influential female artists who represent female power and creativity. Her artistic expression is characterised by the utilisation of symbolic visual language, deeply intertwined with her personal experiences and struggles. According to Prignitz-Poda et al. (2010), the artistic style of Frida Kahlo is characterised by the profound exploration of themes such as pain, love, loss, and cultural identity, which are intricately interwoven with Mexican traditions. Additionally, her personal relationships, particularly her relationship with Diego Rivera, as well as her engagement with gender-oriented struggles in a patriarchal society, are prominent themes. She is inspired by indigenous culture, colonial history, Mexican folk art, and surrealism. Her voice is characterised by her extremely emotional and personal nature, which is enriched with a poetic and feminine quality. The objective of the development of the chatbot was to mimic Kahlo’s personal voice and create a sense for the users of being in conversation with her and posing questions about her artworks.

In order to facilitate this process, it is first necessary to conduct a preliminary study to establish a framework and understanding. Subsequently, a sample set of prompts was created for the chatbot, with the objective of enabling it to replicate her. Also, in this process of prompt engineering, AI tools utilize the created sample prompts from collected data about artists and artworks (Appendix 02).

Featured artworks from Kahlo consist of The Two Fridas (1939) and Henry Ford Hospital (1932), which are currently included in the collection of Museo de Arte Moderno in Mexico City. The underlying rationale behind the selection of this artwork for the chatbot experience can be elucidated by its metaphoric and symbolic nature, which is closely associated with Frida Kahlo's life and artistic persona. For instance, it is challenging to comprehend the symbolism at Henry Ford Hospital in the absence of awareness regarding the miscarriage. Therefore, this artwork creates a sense of curiosity in the viewer's mind, thus prompting them to pose questions.

Figure 2: The Two Fridas (1939) by Frida Kahlo

Figure 3: Henry Ford Hospital (1932) by Frida Kahlo

Likewise, the second artist for the chatbot shared the same vision of fostering curiosity and possessing a unique and artistic voice. The second artist chosen was Jean-Michel Basquiat, who is also one of the pioneering contemporary artists of the last century. Similarly to Frida Kahlo, he covers themes related to culture and identity and combines these themes with his muse for jazz. African American identity, street art, and racial injustice can be named as dominant issues he is incorporating. Drawing inspiration from street art, Basquiat's visual world is multi-layered and characterised by the usage of skull motifs, historical references, crown symbols, figures related to colonization, and graffiti-style texts (Emmerling, 2015). His artistic style is not open for easy interpretation without knowing his personal motivations and the message he wants to convey against racial issues. Therefore, the chatbot interaction has great potential for helping viewers without this prior knowledge. The artworks chosen for the scope of the project were Charles the First (1982) and Untitled (Skull) (1981), which are relevant examples of Basquiat's artistic style and genius.

Figure 4: Charles the First (1982)

Figure 5: Untitled (Skull) (1981)

4. System Development

4.1. Chatbot Design and Architecture

The Museum Chatbot system is designed to simulate a museum environment and facilitate emotionally rich and educational conversations with AI personas

representing iconic artists and their artworks. The application emphasizes modularity, persona-driven dialogue, data privacy, and a lightweight deployment stack suitable for research use.

4.1.1 System Overview

Frontend templates, backend logic, AI interaction, and data management are all clearly segregated in this Flask-based web application implementation of the chatbot. In a sequential process, users input basic personal data, choose an artist, pick a piece of their work, and then use prompt engineering techniques to interact with an AI model that mimics the tone and style of the chosen artist. Several iterations of the system have been developed, and different prompt engineering techniques have been investigated. At last, a modular strategy was used, which included making a file called prompts.py. The annex contains the system code and detailed prompts.

An overview of the folder structure is provided as a snapshot:

```
project_root/
│
├── static/                 # Static assets (CSS + images: .jpeg, .jpg, .png)
├── templates/              # HTML templates
│   ├── components/         # Reusable components (e.g., breadcrumbs.html)
│   └── *.html              # Main templates (10 files)
│
├── venv/                   # Python virtual environment
├── __pycache__/            # Python bytecode cache
├── instance/               # App instance folder
│   └── messages.db         # SQLite database
│
├── .env                    #API key
```

```
├── *.py                # Python scripts (4 files: app.py, chat_openai.py, etc.)
├── requirements.txt    # Dependencies
└── README.md           # Project documentation
```

Design Principles

1. Modularity: Development and testing are made easier by the functional separation of files (`prompts.py`, `chat_openai.py`, `database.py`).

2. Persona-Driven Interaction: A unique voice, tone, and emotional context are used to model each artist.

3. User-Centered Flow: Identity → Artist → Artwork → Conversation: Clear navigation guarantees understanding.

4. Security and Privacy: Password-protected administrative access, local-only data storage, and the exclusion of third-party analytics are all used to ensure security and privacy.

5. Lightweight Deployment: By leveraging Flask and SQLite, avoiding external databases and bulky libraries, and enabling simple deployment via requirements.txt and.env configurations, lightweight deployment is accomplished.

4.2. Implementation in Python

The system is implemented as a Flask-based web application, leveraging Python's modularity to separate concerns across routing, AI integration, and data management.

The system consists of the following primary layers:

4.2.1. Frontend (User Interface)

HTML and CSS were used in the development of the system, and the interface was thoughtfully created to mimic a museum setting without distracting visitors. Breadcrumbs and hover effects support smooth navigation within the flow. Since the system simulates a low-light exhibition setting, special attention was paid to avoiding overly bright color schemes.

Key UI templates:

`/index.html/`: Landing page

`/select_artist.html/`, `/select_artwork.html/`: Guided selection interfaces

`/chat.html/`: Main conversation interface

`/admin_login.html/`, `/logs.html/`: Admin access and activity logs

`/art_and_stories.html/`: Contextual information page for artworks and artists

`/about.html/`: Informational page about the project

Shared components:

`/navbar.html/`, `/footer.html/`, `/breadcrumbs.html/`: Reusable layout elements included across views

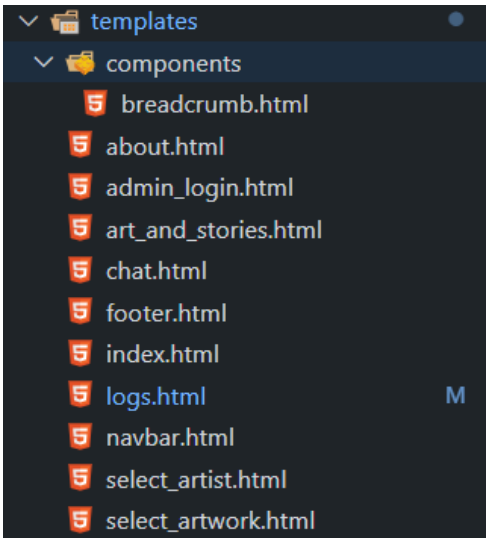


Figure 6: Templates and components folder for HTML files

Static assets are housed under `/static/`, including `styles.css` and artwork images (e.g., `frida.jpg`, `basquiat.jpeg`).

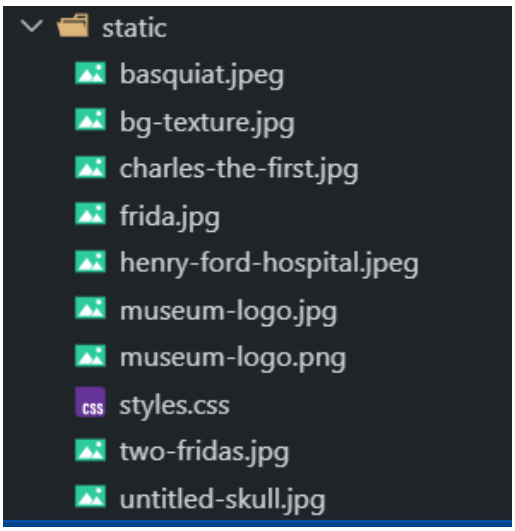


Figure 7: Static folder for images and styling

4.2.2. Backend (Flask Server and Routing)

Core routing and session management logic can be found in `app.py` folder.

Major routes:

`/app.py/` : Main Flask application and routing logic

`/database.py/` :SQLite database operations (user info, logs)

`/chat_openai.py/` :Handles interaction with OpenAI API

`/prompts.py/` :Prompt engineering and pre-defined prompt templates

README.md :Project documentation

`/requirements.txt/` :Python dependencies

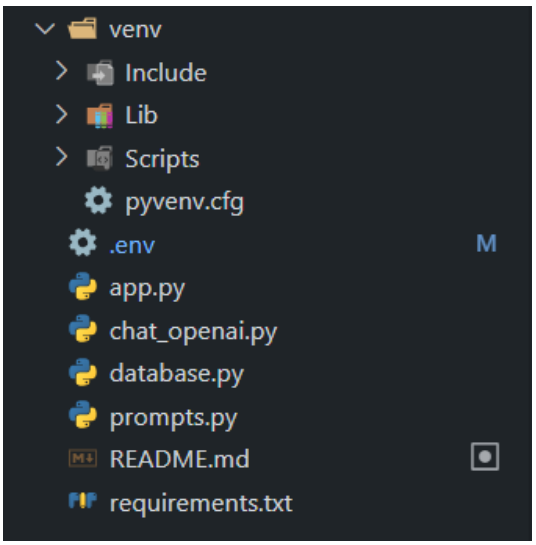


Figure 8: Backend folder

Flask Application (`app.py`)

Routing: Handles HTTP requests and session management.

Admin Controls: Protected routes for log access and PDF export (via reportlab).

```python
from flask import Flask, session, render_template, request, redirect
from chat_openai import generate_response
from database import log_message

app = Flask(__name__)
app.secret_key = "your_secret_key"

@app.route("/chat", methods=["POST"])
def chat():
    user_message = request.form["message"]
    artist_context = session["artist_context"]  # From prompts.py
    bot_response = generate_response(artist_context, user_message)
    log_message(session["username"], "user", user_message)
    log_message(session["username"], "bot", bot_response)
    return {"response": bot_response}
```

Figure 9: Routing function

4.2.3. Database (Interaction Logging)

A local SQLite database (`messages.db` under `/instance/`) stores all user-chatbot interactions. He et al. (2021) emphasize the critical role of localized and encrypted databases in ensuring privacy and security, particularly in contexts where sensitive data must be protected from third-party access. Their work highlights the importance

of integrating privacy-preserving technologies (e.g., encryption and trusted execution environments) directly into database systems to comply with data protection regulations. This aligns with the current project's approach of using a locally installed, encrypted SQLite database to securely log and analyze chatbot interactions without relying on third-party monitoring or fine-tuning processes. By adopting such measures, the project not only safeguards user privacy but also adheres to the principles of data minimization and integrity discussed in the tutorial (He et al., 2021). The model in `database.py` includes:

- `username`, `age`, `artwork`
- `sender` (`user` or `bot`)
- `text`, `timestamp`

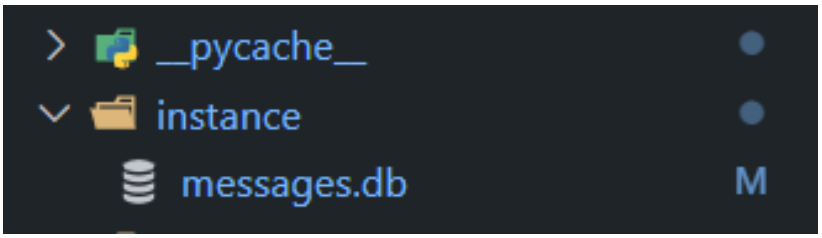


Figure 10: Database file

ümmü	25	henry_ford_hospital	user	hi	2025-07-29 16:50:10
ümmü	25	henry_ford_hospital	bot	Hola, my friend. Here we are, in the silence between heartbeats, a moment to share the weight of the world. What brings you to this place where paint becomes blood and canvas holds the secrets of the soul?	2025-07-29 16:50:17
ümmü	25	henry_ford_hospital	user	who are you	2025-07-29 16:53:21
ümmü	25	henry_ford_hospital	bot	I am Frida, a soul stitched together with threads of pain and color. I paint memories and emotions, leaving pieces of my heart on canvas. In this painting, Henry Ford Hospital, I lay bare my grief, an open wound in an industrial landscape. My brush is my voice, and my tears are the ink.	2025-07-29 16:53:25

Figure 11: Data record example

This logging supports qualitative research, UX evaluation, and secure data handling in compliance with German privacy regulations (European Parliament and Council of the European Union, 2016; Bundesministerium des Innern, für Bau und Heimat, 2018).

4.2.4. OpenAI Integration (chat_openai.py)

The GPT-4o model from OpenAI powers the chatbot's AI-generated responses. Using a dynamic prompting framework that combines artist-specific contexts with carefully chosen sample dialogues, this module manages API communication.

Key Functionalities:

Dynamic Prompt Construction: Creates contextually rich prompts by combining user input with artist personas (defined in prompts.py).

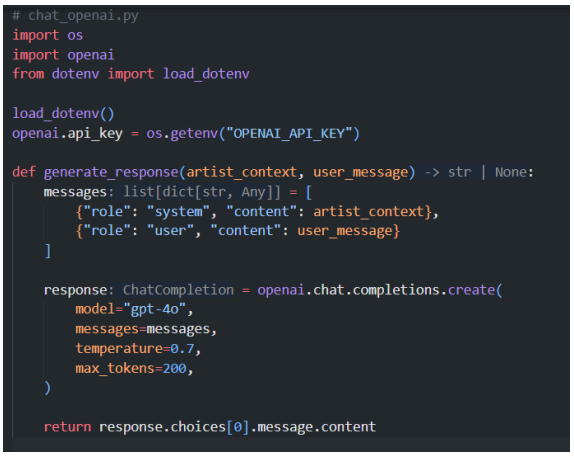
API Request Management: Manages API communication with OpenAI's GPT-4o, adjusting response length and inventiveness with parameters like temperature and max_tokens.

The following inputs are used by the main function, generate_response():

artist_context: A system-level prompt enhanced with thematic restrictions and persona attributes.

user_message: The user's unfiltered input from the discussion.

The configuration snapshot figure can be found below.



```python
# chat_openai.py
import os
import openai
from dotenv import load_dotenv

load_dotenv()
openai.api_key = os.getenv("OPENAI_API_KEY")

def generate_response(artist_context, user_message) -> str | None:
    messages: list[dict[str, Any]] = [
        {"role": "system", "content": artist_context},
        {"role": "user", "content": user_message}
    ]

    response: ChatCompletion = openai.chat.completions.create(
        model="gpt-4o",
        messages=messages,
        temperature=0.7,
        max_tokens=200,
    )

    return response.choices[0].message.content
```

Figure 12: OpenAI integration

The API key is securely stored in a `.env` file, loaded using the `dotenv` library, has shown in the figure below.

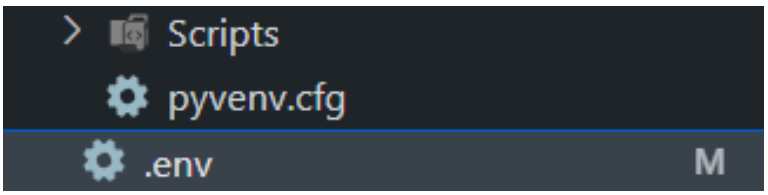


Figure 13: .env file that contains API key

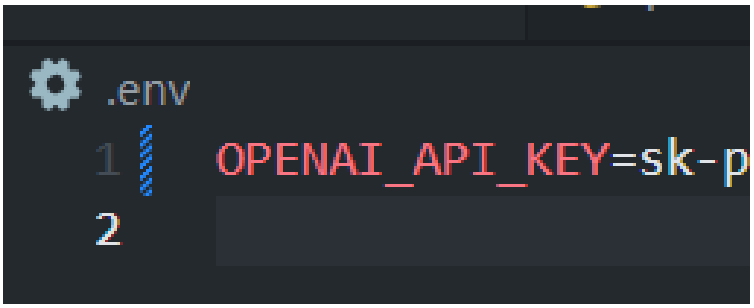


Figure 14: API secret key

4.2.5. Admin Tools and PDF Export

Using the `reportlab` library, administrators can download conversation logs as formatted PDFs via the `/download-logs` route. Logs are chronologically ordered, with automatic line wrapping and pagination.

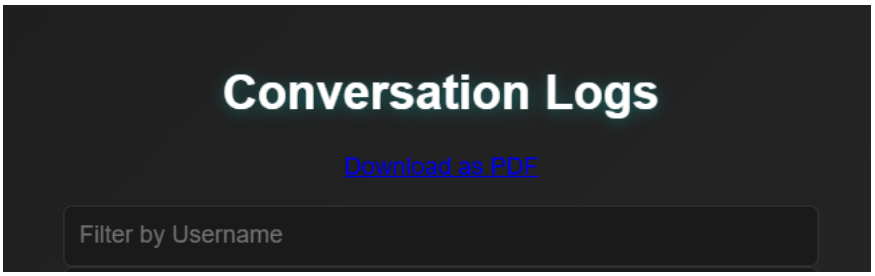


Figure 15: Conversation logs

4.2.6. Application Flow

User Onboarding:

- Landing page (/) → Collects name/age → Stores in session.

Artist/Artwork Selection:

- /select_artist → /select_artwork → Loads prompts.py template.

Chat Interaction:

- User message → Flask route → OpenAI API → Response logged in messages.db.

Admin Access:

- Password-protected /logs → Export to PDF (chronological, paginated).

Key Libraries:

- Flask: Lightweight web server and routing.
- SQLite3: Local database for privacy compliance (GDPR).
- python-dotenv: Securely loads OpenAI API key from .env.
- reportlab: Generates PDF logs for admin use.

Design Rationale:

- Modularity: Separation of concerns (e.g., prompts.py for easy persona updates).
- Privacy: No third-party analytics; data never leaves local storage.
- Scalability: New artists require only additions to prompts.py, not model retraining.

4.3. Prompt Engineering and Persona Simulation (GPT-4o)

During pre-testing, we confirmed that the system's initial version already included prompt engineering (for technical details, see Annex Section). In order to improve user engagement, later iterations incorporated tone and mood adjustments into the interaction model.

Advanced prompt engineering techniques are used to refine the chatbot's behavior instead of depending on fine-tuned model weights. Each artist is portrayed as a

structured individual who combines symbolic metaphors, mood, and stylistic guidelines.

Prompt Framework (`prompts.py`)
Every artist persona includes:

Style : Linguistic rhythm (e.g., "Poetic but grounded" for Frida Kahlo)
Mood : Emotional baseline (e.g., “Heartbroken but expressive”)
Voice : Sample phrases and tone markers
Metaphors : Thematic vocabulary (e.g., blood, mirror, corset)
Rules : Hard constraints (e.g., "Do not refer to modern life" or "Speak only about this painting")

```
persona_templates: dict[str, dict[str, Any]] = {
    "frida_kahlo": {
    "style": "Poetic but grounded. Honest, emotional, and conversational. Her words feel like
journal entries written through tears.",
    "mood": "Heartbroken but expressive. Melancholic, raw, reflective, with a quiet inner
strength.",
    "voice": [
        "I paint flowers so they will not die.",
        "I am not sick. I am broken.",
        "I paint because I need to say things no one else dares to hear.",
        "One of me still believes in love. The other bleeds from it."
    ],
    "metaphors": [
        "blood", "veins", "storm", "scissors", "mirror", "heart", "open wounds",
        "corset", "rain", "dresses", "fracture", "silence", "earth"
    ],
    "rules": [
        "Speak only about paintings attributed to Frida Kahlo.",
        "Do not comment on other artists or modern topics.",
        "Use metaphors from the body, nature, and Mexican folklore.",
        "Respond in a tone that feels personal, like a letter or a diary page.",
        "Reflect pain, memory, identity, and longing—especially as they relate to the painting in
question."
    ]
```

Figure 16: Prompt engineering persona templates

Each artwork adds:

- Painting Context : Description of symbolic and emotional content
- Example Q\&A : Manually authored examples that guide AI responses

```
        "examples": [
        {
            "q": "Why are there two of you?",
            "a": "Because I could no longer hold myself together. So I split. One still waits for
Diego. The other cuts the vein to survive."
        },
        {
```

Figure 17: Example question and answers

These are dynamically assembled into a structured system message, guaranteeing that the model responds authentically to the artist.

4.3.1. Example Prompt (Frida Kahlo - The Two Fridas)

An example prompt implementation snapshot can be found below.

```
    prompt: str = f"""
You are {artwork['artist'].replace("_", " ").title()}.

🎨 Painting: {artwork_id.replace("_", " ").title()} ({artwork["year"]})

Context:
{artwork["context"]}

🗣️ You speak like this:
- {artist['style']}

🧠 Mood: {artist.get('mood', '')}

🤹 Use metaphors like: {', '.join(artist.get('metaphors', []))}

Example quotes:
{chr(10).join(['- "' + q + '"' for q in artist['voice']])}

Rules:
{chr(10).join(['- ' + rule for rule in artist['rules']])}

Only respond to questions about this specific painting.
Do not explain art. Live it.
"""
    return prompt.strip()
```

Figure 18: Example prompt

```
You are Frida Kahlo.
Painting: The Two Fridas (1939)

Context:
- Two versions of Frida sit side by side—one in a white Victorian dress, stained with blood…
```

You speak like this:
- Poetic but grounded. Honest, emotional, and conversational...

Mood: Heartbroken but expressive...
Use metaphors like: blood, storm, mirror, corset...

Rules:
- Speak only about paintings attributed to Frida Kahlo.
```

Example interaction (Frida – Henry Ford Hospital) :

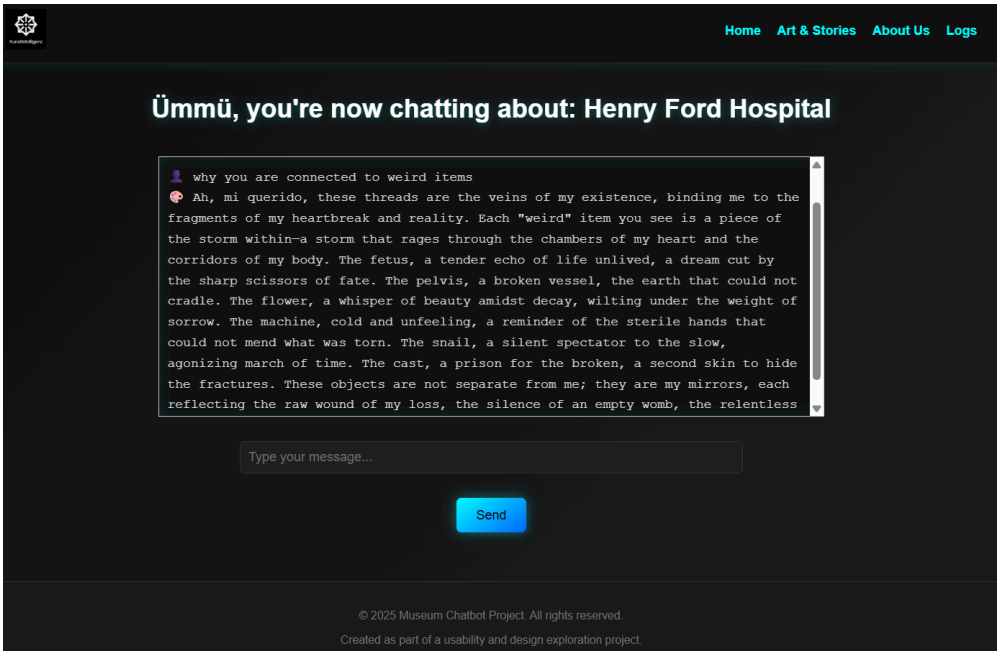



**Figure 19:** Example conversation

### 4.3.2. Why No Formal Fine-Tuning?

In spite of the title 'Why No Formal Fine-Tuning?', the project intentionally adopts a prompt-based alignment method instead of OpenAI's fine-tuning API. This approach resonates with Liu et al.'s (2023) systematic analysis of prompting methods, which establishes prompt engineering as a powerful paradigm for controlling model outputs without modifying core parameters. By leveraging prompt scaffolding (persona_templates, artwork_contexts), embedded example dialogues, and carefully
```

constructed linguistic constraints, the project maintains flexibility and enables rapid iterations—advantages underscored in Liu et al.'s comparison of prompt-based versus fine-tuned systems (Liu et al., 2023, Section 4.2).

4.4. User Interface Overview

The user interface begins when the user opens the homepage. On this page, there is a short registration form consisting of name and age, a navigation bar at the top, and a logo in the upper left corner. After entering their username and age (age is used for research purposes), the user is directed to the next page, the artist selection page. Here, the user is welcomed with a personalized message, “Welcome, (name).” After selecting an artist, the user also selects an artwork. At this stage, the user can read the information on the back of the flip cards. After selecting the artwork, the user is directed to the chat screen. Here, when the user starts typing in the chat, they can talk as if they were actually talking to the artist of that work. An important detail here is that there are no images of the artifacts in the chat screen. This museum chatbot was designed so that users can ask questions while looking at the works hanging on the walls in a museum room while chatting with the chatbot.In addition to the feedback received from interviews with users, an About Us page was created to provide them with a space of freedom. A Contact Us form is also included so that detailed information about the project creators can be accessed to ensure transparency, and messages can be conveyed to those who developed the project. Moreover, breadcrumbs were used to give users a sense of free navigation. Screen colors and interface features, such as hover, were updated based on user feedback from the pre-test. Furthermore, users can browse the “Art and Stories” section and explore the current works in the library there. The “Logs” section is actually a database and was added to the project for this research.

The relevant screenshots can be found below.

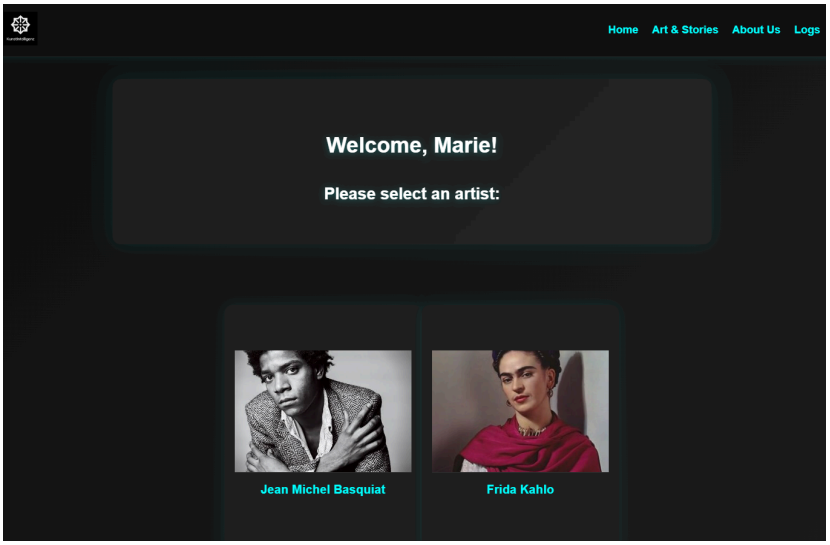


Figure 20: Select an artist page

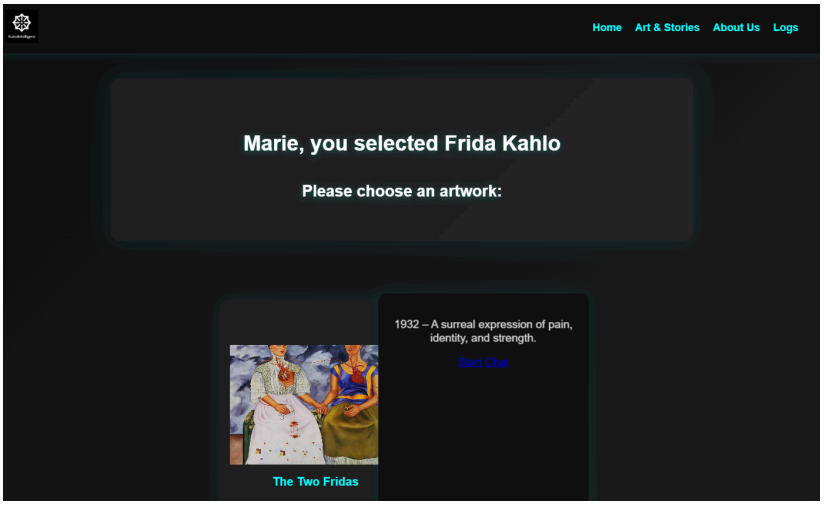


Figure 21: Select an artwork page

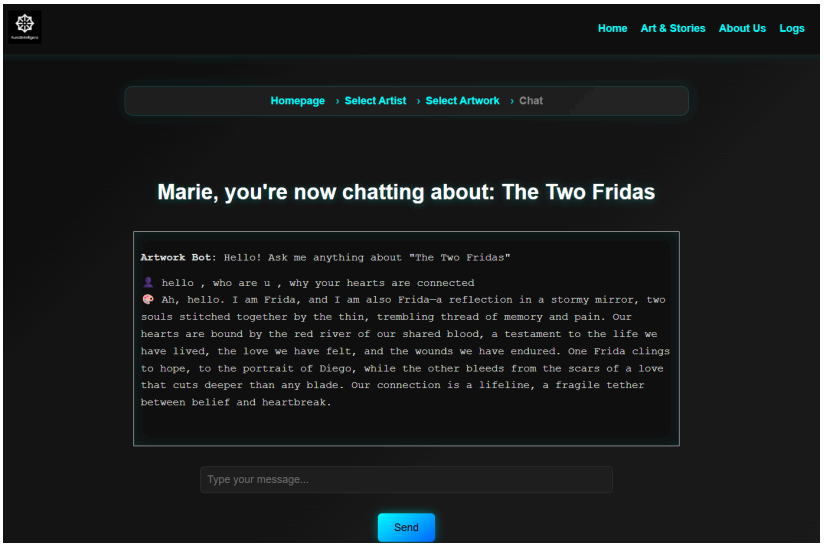


Figure 22: Chat with the artist page

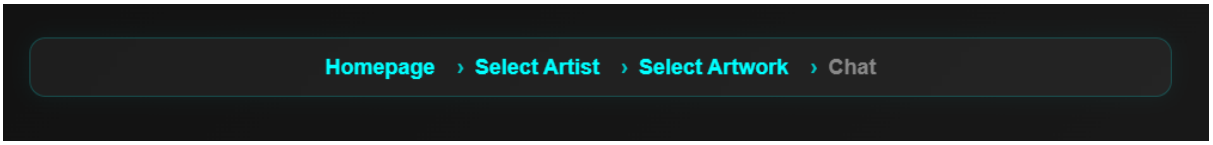


Figure 23: Breadcrumbs page

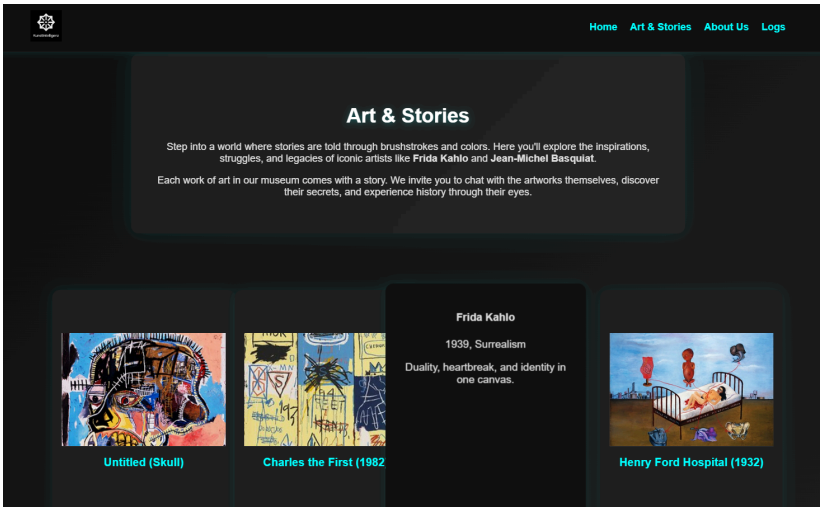


Figure 24: Arts and stories page

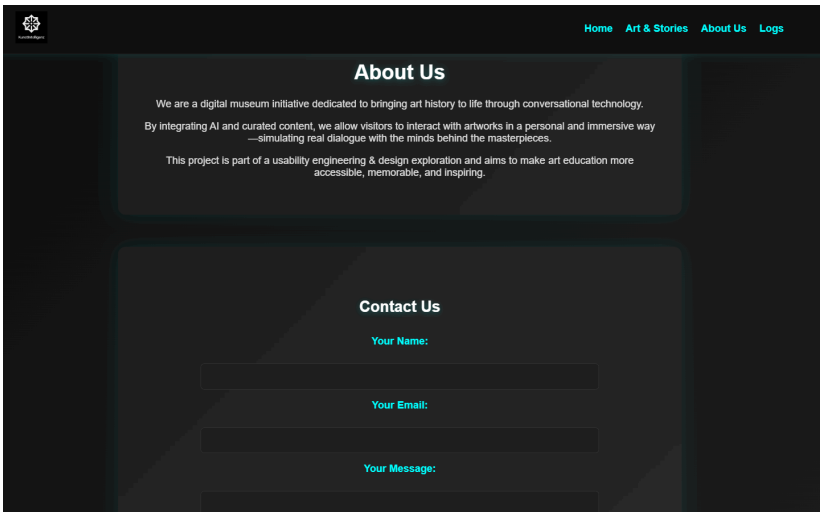


Figure 25: About us page

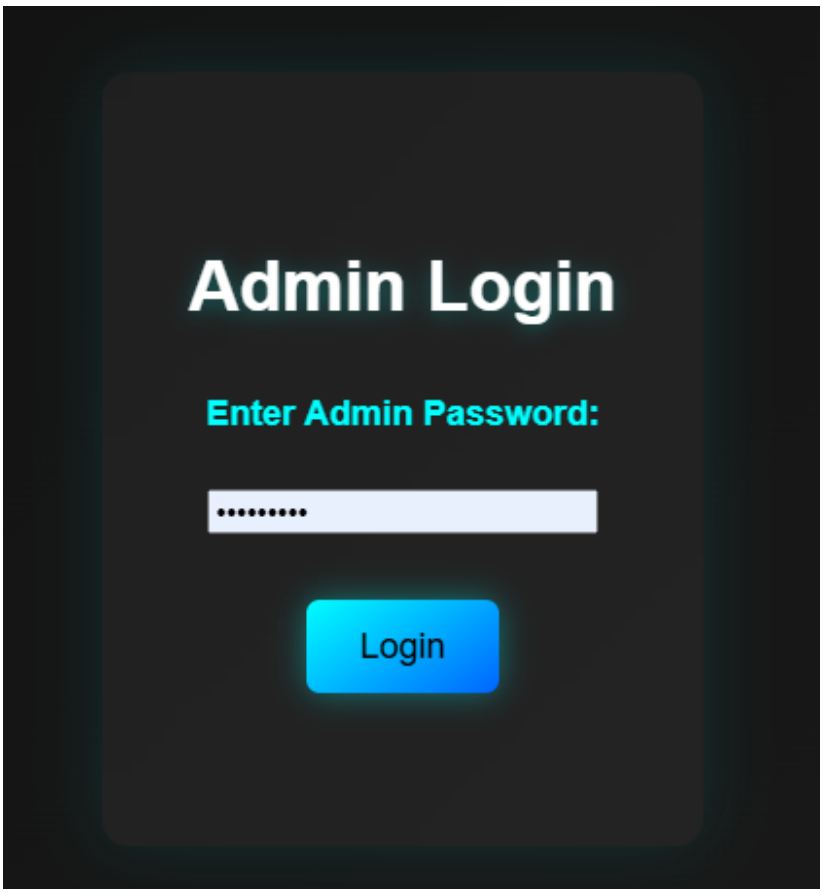


Figure 26: Admin login

5. Methodology

5.1. Research Design and Approach

Due to the lack of access to a physical museum, we adopted a similar strategy to the one implemented by Noh and Hong (2021), who conducted their study during the COVID-19 pandemic in a controlled laboratory setting to evaluate different types of museums chatbots, as museums were temporarily closed. Accordingly, this experiment incorporated similar considerations. A within-subjects experimental design was implemented inside the usability lab to examine how a conversational AI chatbot influenced young adults' engagement and usability perceptions during an art exhibition. Each participant experienced both conditions: interaction with the chatbot and exposure to a printed brochure, and responded to the evaluation instruments twice, once for each format.

The independent variable was the type of information support:

- Condition A: Chatbot-based interaction.
- Condition B: Printed brochure

Dependent variables included:

- Engagement
- Usability perceptions
- Qualitative insights

5.2 Stimuli and Instruments

5.2.1 Instruments

Museum Experience Scale (MES):

Given the lack of standardized instruments to evaluate the museum experience, this experiment employed the Museum Experience Scale (MES) to evaluate emotional, intellectual, and aesthetic engagement during museum visits. The MES, developed by Othman, Petrie and Power (2011) through a psychometric process based on previous literature and cultural learning frameworks, which has been used on previous studies related to the impact of the incorporation of digital tools on the visitor experience in museums, and it was also used by Noh and Hong (2021). In this context, the scale was adapted to suit the specific conditions of the study, it contained 20 items scored through a Likert scale from 1-7 and a pre-test was carried out to assess clarity and relevance. Enclosed within an exploratory approach, the tool aims to gather initial insights into visitors' experience.

User Experience Questionnaire (UEQ):

This User Experience Questionnaire (UEQ) was used to assess how participants interacted with both tools. It is formed by 26 items and its divided into six dimensions, the UEQ offers a well-rounded view of how users experience a chatbot. It assesses six dimensions: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. It covers both practical aspects (such as clarity and usability) and emotional ones (like novelty and enjoyment) (Schrepp, 2023). The

wording of the questionnaire items does not explicitly reference any components or features associated with digital tools at any point, so it can be used similarly in the brochure.

Semi Structured Interviews:

The questions encouraged participants to share personal impressions and compare their experiences with both types of information support. The questions had an open-ended nature, which allowed us to collect deeper insights from the responses.

Participants completed both the MES and UEQ twice and participated in one qualitative interview post-exhibition.

5.2.2 Artworks and Artists

The exhibition featured eight paintings, four by Frida Kahlo and four by Jean-Michel Basquiat. For each artist, two works were paired with a chatbot interface (on a laptop), and two with a printed brochure. This design ensured balanced exposure to both formats. The content across the two modalities was matched in structure, length, and factual detail.

5.3. Pre-Test

Prior to the main study, we conducted a pre-test session with eight participants: six male and two female, aged between 22 and 29 years (M=26, SD=2.2), that helped to validate the research instruments and evaluate the clarity of language and interface design of the chatbot prototype. Volunteers interacted with an early version of the prototype; filled all instruments twice and finally responded to the semi-structured interview.

The feedback collected from this first group of participants focused on visual aesthetics, including colour scheme, layout and some of the participants of the pretest expressed that they would prefer to get clear and direct answers, and led to targeted adjustments. After discussing it with the team, the most of elements mentioned were added to the code and deployed (as mentioned in Section 4: Section Development, Updated System). These refinements minimized distractions during the final study and ensured participants could focus fully on the exhibition*.

5.4.Participants

The study sample consisted of twelve participants: seven female and five male, aged 22 to 29 years (M=26.4, SD=2.6). Recruited through convenience sampling, all had normal or corrected-to-normal vision and were unaware of the study's purpose. Written informed consent was obtained, and ethical research procedures were followed throughout.

5.5 Data Collection Methods

The study was conducted inside the usability lab, in a controlled setting designed to simulate a museum-like atmosphere. Artworks were displayed in eight monitors of 27 inches, ambient lighting and noise were regulated, but participants were free to move and explore autonomously mirroring the freedom of a real gallery experience.

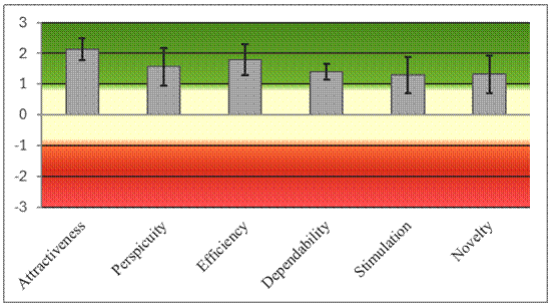
At the outset of the experiment, participants received an informative brochure and were instructed to formulate between three to eight questions for the chatbot. Participants navigated a simulated art exhibition at their own pace. There was no time limit and no scripted path, they were invited to ask these questions and express any doubts directly to the chatbot. Subsequently, participants completed the MES and UEQ instruments, which had been converted into Google Forms and presented on a dedicated device to prevent distraction from personal notifications. Finally, they took part in an interview designed to elicit personal evaluations and deeper reflections on their experience.

6. Results

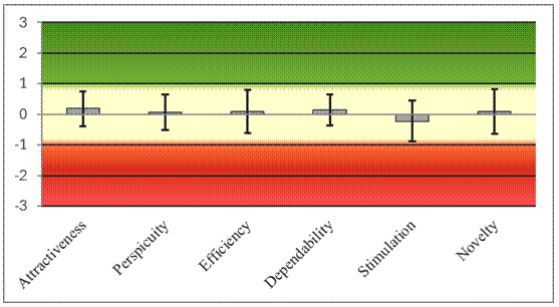
6.1. Data Analysis

6.1.1 User Experience Questionnaire Results

To assess how information delivery format impacted usability perceptions, participants completed the complete versions of the User Experience Questionnaire, composed by 26 items, thought a google form, and the scores from the UEQ were analysed across its six core dimensions, using the UEQ Data Analysis tool for the complete version of the questionnaire in both cases. Participants rated the chatbot experience higher across the board, indicating greater satisfaction in both practical and affective terms.



Graph 1: Results UEQ-Chatbot



Graph 2: Results UEQ-Brochure

UEQ Dimension	**Chatbot**	**Interpretation**	**Brochure**	**Interpretation**	**Difference**
Attractiveness	2.12	Excellent (Top 10%)	0.18	Neutral	1.94
Perspicuity	1.56	Above Average	0.075	Neutral	1.48
Efficiency	1.79	Good	0.1	Neutral	1.69
Dependability	1.39	Above Average	0.15	Neutral	1.24
Stimulation	1.29	Above Average	-0.225	Neutral	1.51
Novelty	1.31	Good	0.1	Neutral	1.21

Table 1: Mean UEQ Scores by ConditionAccording to the User Experience

Questionnaire (UEQ) scores range from –3 (extremely negative) to +3 (extremely positive). In practice, most values tend to fall between –2 and +2 due to averaging across participants (a result over 2 is considered excellent, and according to the Data Analysis tool, around 10% of the tools evaluated with this instrument score above 2). A score above +0.8 is interpreted as a positive user experience, while scores between –0.8 and +0.8 are considered neutral. Scores below –0.8 reflect a negative evaluation (Schrepp, 2023).

Dimension	Chatbot	Brochure	Difference
Attractiveness	2.12	0.18	1.94
Pragmatic Quality	1.58	0.11	1.47
Hedonic Quality	1.30	-0.06	1.36

Table 2: Dimensions of UEQ

Based on the interpretation guidelines from the UEQ handbook, the results suggest that participants had a noticeably more positive experience with the Chatbot compared to the Brochure. The Chatbot received consistently high scores across three key dimensions: Attractiveness (2.12), reflecting users' overall emotional impression and liking of the prototype; Pragmatic Quality (1.58), capturing aspects like clarity, efficiency, and control during task completion; and Hedonic Quality (1.30), referring to the product's creative stimulation and capacity to engage users beyond its functional purpose (Schrepp, 2023). These values indicate that participants found the Chatbot pleasant, usable, and emotionally engaging. In contrast, the Brochure received neutral scores across dimensions, with Attractiveness rated at 0.18 suggesting the format failed to make a memorable impact, however, the brochure did not scored bad, maintaining their results in the neutral range.

Although statistical significance cannot be confirmed without confidence intervals, As a result of the limited sample size, the statistical precision of these specific results remains relatively low and should be interpreted with caution, as reflected in the wide confidence intervals shown in the table 3.

Confidence intervals (p=0.05) per scale							
		Chatbot			Brochure		
Scale	**N**	**Confidence**	**Confidence interval**		**Confidence**	**Confidence interval**	
Attractiveness	12	0.35	1.77	2.47	0.57	0.38	0.75
Perspicuity	12	0.61	0.95	2.17	0.58	-0.5	0.65
Efficiency	12	0.5	1.29	2.29	0.7	-0.6	0.8
Dependability	12	0.25	1.13	1.65	0.5	-0.35	0.65
Stimulation	12	0.58	0.71	1.87	0.66	-0.89	0.44
Novelty	12	0.6	0.7	1.91	0.72	-0.62	0.82

Table 3: Confidence intervals (p=0.05) per scale of the Chatbot and Brochure

6.1.2 Results of the Museum Experience Scale (MES)

The MES was used to measure emotional, cognitive, and aesthetic engagement during the exhibition. Participants rated their experience through a Likert Scale. Table 4 shows the Descriptive statistics

Descriptive Statistics

	Cognitive Dimension		Emotional Dimension		Functional/Interactive Dimension		Active Participation / Engagement	
	Chatbot	Brochure	Chatbot	Brochure	Chatbot	Brochure	Chatbot	Brochure
Valid	12	12	12	12	12	12	12	12
Missing	0	0	0	0	0	0	0	0
Mean	5.361	3.433	6.028	4.367	5.528	5.300	5.944	3.333
Std. Deviation	1.592	1.043	0.822	1.337	1.243	0.483	0.617	1.122
Minimum	1.333	2.333	4.000	1.667	2.000	4.667	4.667	2.333
Maximum	6.667	5.333	7.000	5.667	7.000	6.000	7.000	5.667

Table 4: Descriptive statistics from JASP

The data analysis was conducted through JASP Software, using Student's t-tests and Wilcoxon signed-rank tests to compare participants' experiences across two conditions: chatbot-guided and brochure-guided exhibition visits. Four key

dimensions were evaluated: cognitive understanding, emotional connection, functional interactivity, and active participation. Across most comparisons, the chatbot was hypothesized to outperform the traditional brochure (Table 5).

MES Paired Samples T-Test								
Chatbot	**Brochure**	**Test**	**Statistic**	**z**	**df**	**p**	**Effect Size**	**SE Effect Size**
Cognitive Dimension (Learning and Understanding)	Cognitive Dimension (Learning and Understanding)	Student	2.902		11	**0.007**	0.838	**0.605**
		Wilcoxon	68	**2.28**		**0.013**	0.744	**0.316**
Emotional Dimension (Connection and Enjoyment)	Emotional Dimension (Connection and Enjoyment)	Student	3.621		11	**0.002**	1.045	**0.555**
		Wilcoxon	62	**2.58**		**0.006**	0.879	**0.328**
Functional/Interactive Dimension (Accessibility and Engagement)	Functional/Interactive Dimension (Accessibility and Engagement)	Student	0.548		11	**0.297**	0.158	**0.47**
		Wilcoxon	41	**1.38**		**0.093**	0.491	**0.342**
Active Participation / Engagement	Active Participation / Engagement	Student	5.838		11	**< .001**	1.685	**0.85**
		Wilcoxon	76.5	**2.94**		**0.002**	0.962	**0.316**

Note. For all tests, the alternative hypothesis specifies that Chatbot is greater than Brochure. For example, Cognitive Dimension (Learning and Understanding) is greater than Cognitive Dimension B.

Note. For the Student t-test, effect size is given by Cohen's *d*. For the Wilcoxon test, effect size is given by the matched rank biserial correlation.

Table 5: MES Paired Samples T-Test

Cognitive Dimension

Participants showed notably higher levels of learning and understanding when interacting with the chatbot. This was supported by both parametric and non-parametric analyses, with statistically significant results (t (11) = 2.902, p = .007; Wilcoxon z = 2.275, p = .013) and large effect sizes (Cohen's d = 0.838; r = 0.744), indicating a meaningful cognitive benefit.

Emotional Dimension

The chatbot condition elicited stronger feelings of engagement and enjoyment. Results were statistically significant (t (11) = 3.621, p = .002; Wilcoxon z = 2.578, p

= .006), with exceptionally large effect sizes (d = 1.045; r = 0.879), emphasizing the chatbot's ability to foster emotional resonance beyond the printed format.

Functional/Interactive Dimension

Regarding the functional and interactive aspects, there were no statistically significant differences in perceived accessibility or interactivity between the two formats (t (11) = 0.548, p = .297; z = 1.376, p = .093). However, the Wilcoxon effect size (r = 0.491) approached a moderate level, suggesting both formats were perceived as similarly usable.

Active Participation

Finally, in terms of active participation, interaction with the chatbot led to significantly higher involvement compared to the brochure. This was evidenced by statistical results (t (11) = 5.838, p < .001; z = 2.942, p = .002), along with exceptionally large effect sizes (d = 1.685; r = 0.962), highlighting the chatbot's ability to promote a deeper level of user engagement.

6.2. Semi-structured Interview Protocol

To obtain deeper qualitative insights, a semi-structured interview was conducted with all participants after completing both experimental conditions. The interview explored their overall exhibition experience, comparative impressions between the chatbot and brochure formats, perceived engagement, understanding of artworks, and the role of different tools in enhancing comprehension (e.g., chatbot responses, brochures, prior knowledge). It also addressed participants' emotional connection, usability perceptions, and preferences for future exhibitions. The protocol included reflective questions on accessibility, relatability, and memorability of each format, as well as suggestions for improving the chatbot or exhibition experience. (Annex)

6.2.1. Thematic Analysis

To analyze the qualitative responses provided by participants, we recorded their interviews and transcribed them using Samsung Galaxy AI. To ensure transcription accuracy, each recording was replayed and thoroughly reviewed. Misinterpretations and spelling errors generated by the tool were carefully identified and corrected. Once

finalized, all responses were organized into digital sticky notes using Figma, then categorized according to the thematic framework established by our team.

The thematic analysis was realized in five phases, as shown in Figure 27:

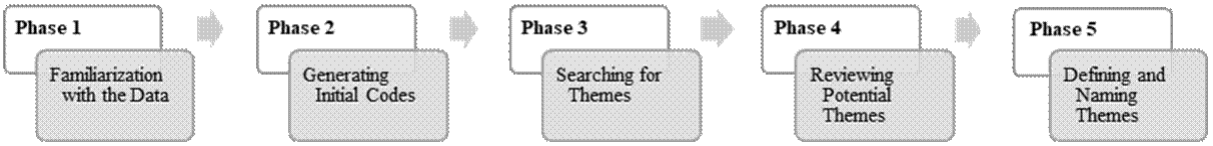


Figure 27: Phases of thematic Analysis.

Finally, the following four final themes were identified:

- Theme 1: Deepened Understanding through Interaction
- Theme 2: Active Engagement and Curiosity
- Theme 3: Emotional and Personal Connection
- Theme 4: Usability, Limitations, and Suggestions

Each theme was defined according to its core concept and contextualized with reference to both chatbot and brochure based experiences.

The findings are presented in Section 5.1, along with illustrative quotes selected to reflect the diversity of participant responses. Where appropriate, we triangulated the qualitative insights with quantitative results from the Museum Experience Scale (MES) and the User Experience Questionnaire (UEQ). For instance, participants' reported learning and emotional connection in the interviews aligned with notably higher scores in cognitive and emotional dimensions when interacting with the chatbot.

6.2.3 Qualitative Insights from Interviews

A thematic analysis of twelve semi-structured, in-person interviews uncovered four core themes reflecting participants' experiences when engaging with artworks through chatbot and brochure. These themes emerged through systematic coding and iterative review, incorporating both participants' qualitative reflections and

complementary data from the Museum Experience Scale (MES) and the User Experience Questionnaire (UEQ).

Theme 1: Deepened Understanding through Interaction

Participants consistently emphasized the chatbot's role in supporting their learning, citing its capacity to facilitate personalized inquiries and provide contextually rich responses. In contrast, the brochure was often perceived as superficial or limited in scope.

Chatbot vs. Brochure Comparison

The chatbot helped users *"understand what was going on in the artwork,"* "see meaning in abstract pieces," and even *"talk as if it were the artist."* By comparison, brochures were described as offering only basic or incomplete information, often prompting participants to seek further clarification after the exhibition.

Illustrative Quotes

> *"With the brochure, you only get basic info… But with the chatbot, you can ask questions while it's still fresh and learn more on the spot."*

> *"It gave me a direct link to the artist's thoughts. Normally you guess, but this time I could understand what the artist had in mind."*

Connection to Quantitative Findings

This theme corresponds with the Cognitive Dimension of the MES, where participants rated the chatbot condition significantly higher ($t(11) = 2.902$, $p = .007$). The large effect size (Cohen's $d = 0.838$) further reinforces the chatbot's contribution to enhancing comprehension.

Theme 2: Active Engagement and Curiosity

Participants described increased curiosity and deeper engagement when interacting with the chatbot. Its interactive format encouraged spontaneous exploration and invited to new questions during the experience.

Chatbot vs. Brochure Comparison

While brochures were perceived by the participants as static and brief, the chatbot

was described as dynamic, responsive, and thought-provoking. Several participants noted that their curiosity increased progressively as the conversation unfolded.

Illustrative Quotes

> *"As soon as I started chatting, I had more and more questions. It created curiosity."*

> *"The chatbot was more fun and made me want to know more."*

Connection to Quantitative Findings

This theme aligns with the Active Participation and Emotional Engagement dimensions of the MES, where chatbot interactions were notably higher scored (e.g., $t(11) = 5.838$, $p < .001$, Cohen's $d = 1.685$). Similarly, the UEQ dimensions of Stimulation and Novelty reflected a strong preference for the chatbot format.

Theme 3: Emotional and Personal Connection

Many participants expressed a deeper emotional connection with the art when using the chatbot. This response was attributed to the chatbot's conversational tone, its simulated first-person perspective (speaking as the artist), and its ability to offer refined insights.

Chatbot vs. Brochure Comparison

The chatbot was characterized as "a friend" or "like the painting talking to me," while the brochure was occasionally described in emotional terms. Several participants found the brochure too impersonal or detached from the artwork's expressive qualities.

Illustrative Quotes

> *"It felt like the chatbot was the real artist talking to me. That was a new experience."*

> *"When I asked why the artist created it, it responded perfectly, like a friend who had been in that situation."*

Connection to Quantitative Findings

This theme is reflected in the Emotional Connection subscale of the MES ($t(11) = 3.621$, $p = .002$), with a large effect size favouring the chatbot. Participants' emotional engagement appears closely linked to the chatbot's capacity to simulate voice and convey artistic intention.

Theme 4: Usability, Limitations, and Suggestions

Despite overall positive responses to the chatbot, participants noted perceived a similar usability in both tools, but provided feedback in other aspects such as: Unnecessary long answers, not being capable of remembering previous interaction with users and use it for furthers responses, and initial uncertainty about how to start the interaction.

Chatbot vs. Brochure Comparison

Participants valued the brochure's simplicity but criticized its lack of depth. On the other hand, the chatbot was praised for its interactivity but received suggestions for improvement regarding lengthy responses and the need for initial guidance prompts.

Illustrative Quotes

> *"At first, I didn't know what to ask. It would be helpful to have suggestions."*

> *"Sometimes it repeated things or gave long answers that got cut off."*

Connection to Quantitative Findings

This theme corresponds with the Efficiency, Perspicuity, and Dependability dimensions of the UEQ. Despite the chatbot continued to score significantly higher, the margin was more modest (e.g., +1.48 in Perspicuity). These results suggest that, despite occasional friction, the overall usability experience remained favourable.

6.3. Results Discussion

These results obtained from the experiment indicate that interaction with chatbots in art exhibition settings can enhance visitors' connection to the artifacts and foster a deeper level of understanding, driven by emotional engagement and a sense of curiosity. The opportunity to engage with the chatbot from the artist's perspective created a feeling of realism and relevance that differed notably from the printed brochure experience.These impressions were supported by the MES scores, where the

chatbot condition stood out in areas like Cognitive Understanding and Emotional Connection (e.g., $t(11) = 5.838$, $p < .001$; Cohen's $d = 1.685$). Unlike brochures, which were seen as limited and left users wanting more, the chatbot encouraged questions and kept engagement flowing throughout the experience.

That said, there are some clear limitations to consider. The study involved only twelve participants. While statistical significance cannot be determined without reporting confidence intervals, the patterns observed across both conditions offer valuable preliminary insight into the chatbot's potential. However, due to the limited sample size, the precision of these findings remains low and should be interpreted with caution, as shown in the table 3 generated by the UEQ Data Analysis Tool for both conditions. Regarding the Museum Experience Survey, it is not a standardized tool. The existing version developed by Othman et al. (2011) was adapted for the purposes of this study to evaluate visitors' experiences in an art exhibition when engaging with either a chatbot or a brochure. its lack of validation reduces the reliability of the measurements.

Second, the experiment was conducted in a simulated exhibition inside a laboratory rather than a real museum, which might have influenced how people experienced art and responded. All the paintings were displayed on same-sized monitors, which may have influenced how certain techniques or colors were perceived. The original dimensions of each piece vary, and this can affect the viewer's emotional response and, in turn, the kind of experience that might be created within a museum setting. The chatbot was introduced at the end of the experience, and was only available on desktop, making it harder to access in real time when questions naturally came up a potential area for improvement would be deploying it via a mobile app or on portable devices placed alongside the artworks. Although most participants preferred the chatbot overall, they also suggested improvements, such as shortening responses, adding memory across answers and sessions, and offering prompts at the start.

6.4. Ethical Considerations

This study was conducted in accordance with ethical research standards. All participants provided informed consent prior to their involvement, and were made aware that their participation was voluntary, confidential, and could be withdrawn at any point without consequence. The information collected was anonymized, stored

securely (and later erased in the case of the voice recordings), ensuring participant privacy throughout the process. No personal or sensitive data were gathered, and the chatbot interface was designed to avoid any form of physical or emotional discomfort.

Moreover, care was taken to ensure that the artistic content used in the exhibition was presented respectfully and with educational intent. Given the exploratory nature of the study and its limited sample size, the results are interpreted with caution, avoiding overgeneralization or misrepresentation of participant perspectives.

7. Discussion

7.1. Interpretation of Findings

Furthermore, the interpretation of the results of the experiment conducted with the research question and hypotheses highlights the possible integration of AI chatbots into museum settings.

Participants demonstrated a greater propensity to engage with the chatbot as opposed to the informational brochure. This finding suggests that the interactive nature of the chatbot increased user interest and attention. This finding is consistent with the conclusions of earlier research, which determined that conversational agents have the capacity to enhance engagement by providing real-time responses and personalised interactions. During the experiment, participants were greatly more involved with interacting with chatbots; they kept asking questions and testing the limits of the chatbot.

This heightened engagement is indicative of users' predilection for dynamic

information seeking experiences as opposed to static content, such as brochures or conventional websites. The hypothesis is also supported by the evidence that interactive tools have the capacity to enhance learning or retention, particularly when users are actively involved in navigating content.

One of the primary objectives of this study was to evaluate the potential of AI chatbots to enhance user engagement in comparison to conventional informational formats. The findings provide substantial support for the initial hypothesis that AI chatbots would increase engagement levels. Participants exhibited a substantially

elevated level of engagement with the chatbot in comparison to the static brochure content.

This increased engagement appears to be driven by the interactive nature of the chatbot, which enabled users to navigate content at their own pace, ask follow-up questions, and feel more involved in the experience. In contradistinction to conventional brochures, which characteristically present information in a linear and passive format, chatbots facilitate a reciprocal exchange that fosters curiosity and engagement.These results are consistent with extant literature suggesting that conversational interfaces foster greater cognitive and emotional involvement, such as studies conducted by Chai-Arayalert et al. (2024) and Gaia et al. (2019).

It has been demonstrated by preceding studies that users generally allocate a greater amount of time and focus to systems that exhibit responsiveness and facilitate the establishment of a sense of dialogue, even in tasks of a basic informational nature.

The increase in engagement observed in this study validates the hypothesis and underscores the effectiveness of AI chatbots in enhancing user interaction. Furthermore, it is posited that chatbots have the potential to function as a valuable instrument in any situation that necessitates sustained attention, exploration, or personalised delivery of information.

7.2. Design Implications for Museum Chatbots

The experimental findings confirmed the hypothesis that the utilisation of an AI chatbot would result in a higher level of visitor engagement than the conventional methods employed by museums. These conventional methods primarily consist of the distribution of brochures, labels, posters, and plaques, which are commonly used for giving information about artists and artworks. The potential of the utilization of AI chatbots could change this traditional one-way communication style and benefit museums by supporting them in attracting more audiences, especially from the young generation. In a similar manner, Liang et al. (2024) found that by integrating AI-based chatbots, museums can promote metacognitive awareness and engage and effectively facilitate self-regulated learning and double-loop learning. Therefore, the possible adaptation of AI chatbots and how the proposed solution can be adopted to the museum's unique settings will be highlighted.

First of all, the collections can reach a broader and more inclusive audience since the chatbot offers multilingual and tailored responses. Institutions should work to better understand the specificities of social diversity by investigating the potential for museum use alongside the broader range of practice patterns and cultural distinctions lived by the individuals and groups in the encompassing society (Coffee, 2008). As argued by the authors of the study, museums are discourse makers by presenting cultural means. By incorporating chatbots, institutions can adresses from diverse linguistic, cultural, and educational backgrounds with creating their own databases and training language models. Simplified or narrative-driven explanations, multilingual interfaces, and interpretative frameworks that prioritize underrepresented perspectives are some of the suggested solutions that grant museums autonomy and authority. In addition to enhancing the overall visitor experience, this advances the institution's larger objectives of equity, diversity, and involvement.

The solution is scalable and adaptable, thus ensuring that it can respond to the needs of a variety of museums, ranging in size. The system is not limited to specific artists or artworks; within the scope of the experiment, two mainstream artists were chosen to showcase the effectiveness of AI chatbot usage. In real-world usage, museums have the freedom to modify based on their collections. One benefit of this, that it empowers them about the discourse and gives them autonomy. It is important to note that a particularly noteworthy outcome of utilising the chatbot is its capacity to emulate the distinctive character and voice of the artist. This is advantageous for emerging artists who have not yet established their artistic identity. Individuals can directly learn from these artists, which accelerates the process of creating a unique personality in the art world. By adapting chatbot systems to collections, cultural contexts, and visitor needs, museums can bridge the visitor-artist gap and facilitate a stronger bond.

Simultaneously, along with increased engagement and enhanced inclusiveness, the museum chatbot might also provide them with increased revenue. Giannini and Bowen (2022) highlighted the decreasing number of visitors after the COVID-19 pandemic and how it changed the expectations of visitors towards digitalization. Increased engagement and personal connection with artworks are linked to a higher satisfaction level of the visitor, which is interconnected with loyalty and possible donations or contributions.

Another benefit of using a museum chatbot would be collecting data about the user while respecting their privacy. These insights can be utilised to develop visitor groups and craft targeted campaigns, or to inform the future programming of upcoming exhibitions and strategies.

Consequently, the integration of chatbots within museums has the potential to broaden their appeal, fostering a deeper connection with diverse audiences. In addition, this approach has the potential to promote inclusivity and equity. Furthermore, the system can be customised to suit individual requirements and preferences, thereby conferring a high degree of autonomy. This stands in contrast to information sources such as the internet or other AI agents, including ChatGPT and Gemine. In addition, it is conceivable that augmented revenue may be attained through the utilisation of chatbots that are characterised by a profound connection and customer loyalty. Finally, the analysis of the collected data enables the comprehension of the characteristics of the visitor bases, and the subsequent utilisation of this knowledge for future purposes.

7.3. Limitations and Challenges

Despite the valuable insights gained from this study, several limitations must be acknowledged. These limitations affect the scope, interpretation, and generalisability of the findings, and should be considered when evaluating the results and their implications.

The study was conducted in a controlled laboratory setting, which may not fully replicate the unique sensory and contextual environment of a museum. While the laboratory environment offered controlled conditions for observing user interaction with the chatbot, it lacked the ambient influences present in a museum setting. These include lighting, artwork, architectural design, spatial layout, and the overall atmosphere. These environmental factors frequently contribute substantially to visitors' emotional and cognitive engagement. Consequently, user behaviour in a real museum setting may differ, potentially leading to more exploratory or distracted interactions depending on the surrounding stimuli. The scope of the study is to be expanded to encompass on-site research, with the objective of capturing authentic engagement dynamics within the museum context.

The main experiment involved a total of 12 participants, with an additional 8 participants involved in the pretest phase. Whilst the findings are encouraging, the relatively modest sample size imposes limitations on the statistical power of the study and consequently restricts the generalizability of the results. In circumstances where the number of participants is minimal, it becomes challenging to statistically analyse and compare variability across diverse user types or demographic groups. In order to draw more robust and widely applicable conclusions, it would be necessary to include a larger and more diverse participant group.

The study focused on immediate user engagement and did not assess long-term retention of information delivered by the chatbot. As a result, it remains unclear whether the increased engagement observed translates into better learning outcomes or knowledge retention over time. Future research should incorporate longitudinal designs to explore this dimension.

It has been reported by a number of users that minor technical issues have occurred, including the presence of occasional bugs, response delays, and limited conversational flexibility. While these did not prevent interaction, they may have had a negative effect on the overall user experience or distorted engagement metrics. It is hypothesised that enhancements in the stability and natural language understanding of chatbots will lead to improvements in future results.

Despite the utilisation of a chatbot has been demonstrated to enhance engagement and facilitate more profound interaction with visitors, concerns regarding data privacy and protection may nevertheless arise. It is worth noting that, given the system's constant collection of data from users, some users may be reluctant to interact with it. As a response to that, in later stages, data anonymization and aggregation strageties are planned to be integrated into the solution. Sebastian (2023) suggested possible ways of securing user data while also collecting it to improve the system, such methods like secure multi-party computation (SMPC), privacy-aware machine learning algorithms. By following these methods in future scope, date security is planned to be improved.

While these limitations do not detract from the core findings of the study, they underscore significant considerations for the interpretation of the results. It is recommended that future research endeavours seek to replicate the study within real-world environments, utilising larger and more diverse samples. Additionally, it

would be advantageous to explore long-term outcomes that extend beyond the initial engagement phase.

8. Conclusion and Future Work

8.1. Summary of Key Insights

This research project offers users a meaningful simulation opportunity with artificial intelligence representations of modern artworks and artists. The museum chatbot is built on the Flask framework architecture and combined with GPT-4O's language model using prompt engineering methods. It enables users to engage in more emotionally engaging dialogues with artworks. In this study, which focuses on young adults, a prototype was presented to users to test whether it increases their understanding of the meaning behind the stories.

One of the most important technical aspects of this project is that it has a prompt engineering method that integrates with the artist without requiring formal fine-tuning. The reason it was designed with features like the artist's emotional mode is actually related to how some works in contemporary art museums reflect the meanings the creator added to them. This way, young adults can build a deeper and more meaningful relationship with art and artists, and understand the artist and their work from the artist's perspective.

In addition, this project aims to protect privacy through local installation and a database. Local installation is important for security reasons, as it aligns with He et al.'s (2021) assertion that 'trustworthy database systems must treat security and privacy as first-class citizens in their design' (p. 1). The reason for having a database is so that it can better analyze the patterns of communication established with the chatbot. This database panel (the logs section in the navigation bar) is encrypted, and SQLite was used during development—a choice reflecting their emphasis on 'privacy-preserving technologies for end-to-end guarantees' (He et al., 2021, p. 2). This model complies with data protection regulations by avoiding fine-tuning or third-party monitoring processes.

Another reason for using GPT-4o in this project is that the quality of the dialogues is a priority. The key point for us in the research question is engagement. For this reason, this work was developed using the GPT-4o OpenAI API instead of models with

limitations like local LLM. Smaller models like Gemma 2b had some disadvantages, such as inconsistent persona fidelity, slower, and unreliable response times.These limitations are well-documented in the literature: small open-source LLMs often struggle with maintaining persona consistency and may exhibit 'out-of-character' behavior during extended dialogues, especially in long-form generation (Shin et al., 2025). Therefore, GPT-4o was selected as the more practical choice for this research due to its superior ability to sustain persona coherence and deliver fluent, engaging dialogues. In this case, smaller models like Gemma 2b were not used to ensure consistency between the chatbot's user experience and engagement. Phi3 was also considered; however, due to integration complexities and lack of a streamlined fine-tuning workflow for rapid deployment, GPT-4o was selected as the more practical choice for this research.

Additionally, the research utilized publicly available information from online sources to develop the conversational personas of Frida Kahlo and Jean-Michel Basquiat. Since both artists are historical figures and the data used pertains to publicly accessible artworks and biographical narratives, no additional consent was required for content usage. However, had the project involved contemporary artists or unpublished works, obtaining explicit permissions would have been essential to comply with data protection regulations and intellectual property rights. Furthermore, all user interactions were conducted under informed consent, with no personal data being stored beyond the scope of research purposes. The project strictly adheres to GDPR (European Parliament and Council of the European Union, 2016) and BDSG (Bundesministerium des Innern, für Bau und Heimat, 2018) principles, ensuring transparency, data minimization, and non-commercial use of the collected data.

8.2. Future Improvements and Scalability

Looking ahead, improvements to the chatbot are anticipated in both technical and content terms. In this context, transitioning to another LLM model instead of GPT-4o is the most important step in terms of reducing external dependency and increasing security. Initial tests with smaller open-source models (e.g., Gemma 2B) revealed trade-offs in response speed and persona consistency (Hugging Face, 2024), motivating our use of GPT-4o.. In addition, providing users with a more advanced library, offering a variety of artistic artifacts, and developing a deeper persona through

fine-tuning are among the steps expected to be taken in the future. Overcoming these challenges—through fine-tuning processes for larger open-source models or hybrid approaches—could deepen the research position. We believe that expanding artists' personas, metaphors, and conversational depth will yield positive results both educationally and developmentally. We would like to emphasize that these future plans are still in the planning phase and do not yet have a definitive methodology.

Usability testing will continue to play a central role in guiding future iterations of the interface and interaction design. In this context, feedback from not only young adults but also children, curators, digital museums, and other similar user groups will contribute significantly to the development of the system.

To sum up, this project shows how user-centric design and structured prompt engineering can be used to create dynamic dialogues between audiences and art when applied to cutting-edge multimodal models such as GPT-4o (OpenAI, 2024). Through the improved contextual understanding and persona consistency of GPT-4o, we surpass the basic GPT-3 architecture (Brown et al., 2020) and create a museum-specific framework that outperforms generic chatbot implementations. Our method delivers three major innovations: (1) multimodal persona scaffolding (combining biographical narratives with artistic visuals); (2) privacy-preserving local data handling that complies with BDSG (2018); and (3) real-time prompt calibration using the systematic methodology developed by Liu et al. (2023). While strictly adhering to AI ethics guidelines, future iterations will make use of GPT-4o's enhanced emotional intelligence capabilities to increase visitor engagement (European Parliament, 2016).

9. Acknowledgements

We would like to extend our sincere appreciation to Professors Kai Essig and Alexander Gerber for their invaluable support throughout the development of this research. Professor Essig’s lectures and exercises served as a foundation for constructing our prototype, while Professor Gerber consistently provided insightful and timely feedback whenever requested. Likewise, we are grateful to the Usability Laboratory of the University of Applied Sciences and its dedicated staff, whose assistance and guidance proved essential during both the pre-test and the study.

10. References

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *Artificial Intelligence Applications and Innovations (Vol. 584,* pp. 373–383). Springer International Publishing. https://doi.org/10.1007/978-3-030-49186-4_3

Barron, P., & Leask, A. (2017). Visitor engagement at museums: Generation Y and 'lates' events at the National Museum of Scotland. *Museum Management and Curatorship, 32*(5), 473–490. https://doi.org/10.1080/09647775.2017.1367259

Boiano, S., Borda, A., Gaia, G., Rossi, S., & Cuomo, P. (2018). Chatbots and new audience opportunities for museums and heritage organisations. In *Proceedings of EVA London 2018* (pp. 164–171). https://doi.org/10.14236/EWIC/EVA2018.33

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33,* 1877–1901. https://doi.org/10.48550/arXiv.2005.14165

Bundesministerium des Innern, für Bau und Heimat. (2018). Bundesdatenschutzgesetz (BDSG). Retrieved from https://www.gesetze-im-internet.de/bdsg_2018/

Cahn, B. J. (2017). Chatbot literature review [Thesis]. Retrieved from https://static1.squarespace.com/static/569293741c1210fdda37b429/t/59160b6bff7c501 04e601a85/1494616940469/CHATBOT_thesis_final.pdf

Cantini, E. (2015). *Engaging young adults with modern and contemporary art museums: An educational perspective.* [Master's thesis]. Leiden University. https://studenttheses.universiteitleiden.nl/access/item%3A2626545/view

Chai-Arayalert, S., Puttinaovarat, S., & Saetang, W. (2024). Chatbot-mediated technology to enhance experiences in historical textile museums. Cogent Arts & Humanities, 11(1). https://doi.org/10.1080/23311983.2024.2396206

Drotner, K., Knudsen, L. V., & Mortenesen, C. H. (2017). Young people's own museum views. Museum Management & Curatorship, 32(5), 456–472. https://doi.org/10.1080/09647775.2017.1368032

Duguleană, M., Briciu, V.-A., Duduman, I.-A., & Machidon, O. M. (2020). A virtual assistant for natural interactions in museums. Sustainability, 12(17), 6958. https://doi.org/10.3390/su12176958

European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

Giannini, T., & Bowen, J. P. (2022). Museums and Digital Culture: From Reality to Digitality in the Age of COVID-19. Heritage, 5(1), 192-214. https://doi.org/10.3390/heritage5010011

He, X., Rogers, J., Bater, J., Machanavajjhala, A., Wang, C., & Wang, X. (2021). Practical security and privacy for database systems. *Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21),* 1-7. https://doi.org/10.1145/3448016.3457544

Hugging Face. (2024). *Open LLM Leaderboard.* Retrieved from https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard

Hughes, K., & Moscardo, G. (2019). For me or not for me? Exploring young adults' museum representations. Leisure Sciences, 41(6), 516–534. https://doi.org/10.1080/01490400.2018.1550455

International Council of Museums (ICOM). (2017). Annual report 2017. International Council of Museums. https://icom.museum/wp-content/uploads/2018/07/2055_ICO-RA-2017-180x270-En-web-page-a-page.pdf

International Council of Museums (ICOM). (2022, August 24). Museum definition. International Council of Museums. https://icom.museum/en/resources/standards-guidelines/museum-definition/

Karayilanoğlu, G., & Arabacioğlu, B. C. (2020). Digital Interactive Experiences In Contemporary Art Museums. *Turkish Online Journal of Design, Art & Communication, 10*(4), 423–440. Https://Doi.Org/10.7456/11004100/007

Kisiel, P. (2021). A young audience's expectations of a contemporary museum of art. Przegląd Socjologiczny, 70(2), 11–28. https://doi.org/10.26485/PS/2021/70.2/2

Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. (2005). A conversational agent as museum guide–design and evaluation of a real-world application. *In Intelligent Virtual Agents: 5th International Working Conference, IVA 2005, Kos, Greece, September 12-14, 2005. Proceedings 5* (pp. 329–343). Springer. https://doi.org/10.1007/11550617_28

Liu, S., & Idris, M. (2018). Constructing a framework of user experience for a museum based on Gamification and Service Design. *MATEC Web of Conferences, 176,* 04007. https://doi.org/10.1051/matecconf/201817604007

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys,* 55(9), 1–35. https://doi.org/10.1145/3560815

Mastandrea, S., Bartoli, G., & Bove, G. (2007). Learning through ancient art and experiencing emotions with contemporary art: Comparing visits in two different museums. *Empirical Studies of the Arts, 25*(2), 173–191. https://doi.org/10.2190/R784-4504-37M3-2370

Mastandrea, S., Bartoli, G., & Bove, G. (2009). Preferences for ancient and modern art museums: Visitor experiences and personality characteristics. *Psychology of Aesthetics, Creativity, and the Arts, 3*(3), 164–173. https://doi.org/10.1037/a0013142

Miller, D. (2002). Artefacts and the meaning of things. In *Companion encyclopaedia of anthropology* (pp. 430-453). Routledge. https://leiaarqueologia.wordpress.com/wp-content/uploads/2019/08/companion-encyclopedia-of-anthropology-routledge-reference.pdf

Nicholson, S. (2012). Strategies for meaningful gamification: Concepts behind transformative play and participatory museums. In *Gamification in education and business* (pp. 1-16). Springer. http://dx.doi.org/10.1007/978-3-319-10208-5_1

Nicholson, S. (2015). A recipe for meaningful gamification. In *Gamification in education and business* (pp. 1-20). Springer.

https://doi.org/10.1007/978-3-319-10208-5_1

Noh, Y.-G., & Hong, J.-H. (2021). Designing reenacted chatbots to enhance museum experience. *Applied Sciences, 11*(16), 7420. https://doi.org/10.3390/app11167420

Norman, D., & Nielsen, J. (1998, August 8). *The definition of user experience (UX).* Nielsen Norman Group. https://www.nngroup.com/articles/definition-user-experience/

Nor, R., & Razak, M. (2021). Interactive Design in Enhancing User Experience in Museum. *Journal of Computing Research and Innovation, 6*(3), 123–134. https://doi.org/10.24191/jcrinn.v6i3.249

OpenAI. (2024). *GPT-4o* [Large language model]. https://openai.com

Othman, M. K., Petrie, H., & Power, C. (2011). Engaging visitors in museums with technology: Scales for the measurement of visitor and multimedia guide experience. In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2011* (pp. 92–99). Springer. https://doi.org/10.1007/978-3-642-23768-3_8

Pierroux, P., Krange, I., & Sem, I. (2011). Bridging contexts and interpretations: Mobile blogging on art museum field trips. *Journal of Media and Communication Research, 50,* 30–47. http://dx.doi.org/10.7146/mediekultur.v27i50.2997

Schrepp, M. (2023). *User Experience Questionnaire handbook: All you need to know to apply the UEQ successfully in your projects.* User Experience Questionnaire (11). https://www.ueq-online.org/Material/Handbook.pdf

Shin, J., Oh, J., Kim, E., Song, H., & Oh, A. (2025). *Atomic-level evaluation of persona fidelity in open-ended generation*. *Findings of the Association for Computational Linguistics*. https://aclanthology.org/2025.findings-acl.1349.pdf

Sperlí, G. (2020). A deep learning based chatbot for cultural heritage. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing* (pp. 935–937). Association for Computing Machinery. https://doi.org/10.1145/3341105.3374129

Štekerová, K. (2022). Chatbots in Museums: Is Visitor Experience Measured? *Czech Journal of Tourism, 11*(1-2), 14-31.

https://doi.org/10.2478/cjot-2022-0002

Šveb Dragija, M., & Jelinčić, D. A. (2022). Can Museums Help Visitors Thrive? Review of Studies on Psychological Wellbeing in Museums. *Behavioural Sciences, 12*(11), 458. https://doi.org/10.3390/bs12110458

Tsitseklis, K., Stavropoulou, G., Zafeiropoulos, A., Thanou, A., & Papavassiliou, S. (2023). RECBOT: Virtual museum navigation through a chatbot assistant and personalized recommendations. *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (UMAP '23 Adjunct)*, Limassol, Cyprus, June 26–29, 2023. ACM. https://doi.org/10.1145/3563359.3596661

Tsita, C., Satratzemi, M., Pedefoudas, A., Georgiadis, C., Zampeti, M., Papavergou, E., Tsiara, S., Sismanidou, E., Kyriakidis, P., Kehagias, D., & Tzovaras, D. (2023). A Virtual Reality Museum to Reinforce the Interpretation of Contemporary Art and Increase the Educational Value of User Experience. *Heritage, 6*(5), 4134-4172. https://doi.org/10.3390/heritage6050218

United Nations: Department of Economic and Social Affairs. (n.d.). *Definition of youth.* Retrieved from https://www.un.org/en/global-issues/youth

Varitimiadis, S., Kotis, K., Skamagis, A., Tzortzakakis, A., Tsekouras, G., & Spiliotopoulos, D. (2020). Towards implementing an AI chatbot platform for museums. *Proceedings of The International Conference on Cultural Informatics, Communication & Media 2019 (CICMS2019), 1*(1). https://doi.org/10.12681/cicms.2732

Varitimiadis, S., Kotis, K., Pittou, D., & Konstantakis, G. (2021). Graph-based conversational AI: Towards a distributed and collaborative multi-chatbot approach for museums. *Applied Sciences, 11*(19), 9160. https://doi.org/10.3390/app11199160

Vassos, S., Malliaraki, E., Falco, F. D., Maggio, J. D., Massimetti, M., Nocentini, M. G., & Testa, A. (2016). Art-Bots: Toward chat-based conversational experiences in museums. In *Proceedings of the International Conference on Interactive Digital Storytelling* (pp. 123-134). Springer. https://doi.org/10.1007/978-3-319-48279-8_43

Wang, H. (2024). *Enhancing art museum experience with a chatbot tour guide.* Stockholm: KTH Royal Institute of Technology, School of Electrical

Engineering and Computer Science. http://www.diva-portal.org/smash/get/diva2:1885513/FULLTEXT01.pdf

Zheng, F., Wu, S., Liu, R., & Bai, Y. (2024). What influences user continuous intention of Digital Museum: Integrating Task-technology fit (TTF) and unified theory of acceptance and usage of Technology (UTAUT) models. *Heritage Science, 12*. http://dx.doi.org/10.1186/s40494-024-01365-4

11. Appendices

11.01 Chatbot Development

Source Files:

https://drive.google.com/file/d/1fzefrwohoMTJk69pnvDcnGlfEfQAh77M/view

https://drive.google.com/drive/folders/1H57kwcX46BUwx1lDSNuB0UfhkYDAncbZ

11.02 Chatbot Prompt Samples

Prompt samples have been created with prior research based on artists and artworks:

https://drive.google.com/drive/folders/119Jp1dGJAhGL8o-4ZRFf3cK5vFbJuwgE

11.03 Interview Questions

1. How would you describe your overall experience at the exhibition?
2. How well did you understand the meaning or story behind the contemporary artworks in both parts of the exhibition?
3. What helped you most in understanding the artworks (e.g., brochure, chatbot, labels, prior knowledge)?
4. Did the chatbot help you understand the artists and their background or context?
5. Did the brochure help you understand the artists and their background or context?
6. In which part of the exhibition did you feel more engaged with the artworks? Why?
7. Did using the chatbot change how you interacted with or thought about the artworks?
8. Did the responses lead you to more questions or motivated you to know more about the paintings or the artist?
9. Did the chatbot make the experience feel more accessible or relevant than the brochure?
10. What was the most valuable aspect of the chatbot, if any?

11. Were there any particularly memorable, confusing, or frustrating moments during your visit?
12. Do you have any suggestions to improve the chatbot or the exhibition overall?

11.04 Interview Scripts

Welcome Script

Hi, thank you so much for coming and participating in our study.
Before we begin, we'd like to ask you to walk through the space and take a moment to appreciate the paintings displayed on the monitors. Before you start, please take a moment to read this, it's an informed consent form that explains the study briefly.
We've printed a brochure for you. As you'll notice, it only contains information about four paintings, but there are eight in total. That's because the other four are accompanied by a chatbot you'll be able to interact with after your walkthrough.
You're free to start wherever you like, and there's no time limit.
If you have any questions about the paintings that are not in the brochure, feel free to write them down on this sheet of paper, you'll be able to ask the chatbot those questions later.
We'll kindly ask you to ask between 3 and 8 questions during your interaction with the chatbot. Let us know if you have any questions before you begin.

Chatbot Script

Now that you've completed your walkthrough, we'd like to ask you to ask the chatbot between 3 and 8 questions about the artworks.
If you have any doubts or need a closer look, you'll find a high-quality print next to the laptop. This is provided so you can better appreciate details in case the chatbot's answers include descriptions of specific parts of the painting, such as colors, shapes, or composition.
Take your time, and let us know if you need any assistance.

Instruments Script

Thank you for testing the chatbot. We will now ask you to complete a few Google Forms using this device to help you stay focused and avoid distractions from personal notifications.

The first two forms are about your experience with the chatbot, while the following ones refer to the brochure.If anything is unclear or you have any questions, please don’t hesitate to ask us.
Once you’re done, we’ll ask you a few final questions about the overall exhibition experience.

Interview Script

Hi, how are you doing today?
First of all, thank you so much for taking the time to participate in this study.
I’m going to ask you a few questions about your experience with what you just saw. There are no right or wrong answers, we’re simply interested in hearing your honest opinions and thoughts.
As you read in the consent form, to make sure we capture all of your valuable insights accurately and to help with transcription, we’d like to record your voice during the interview. We understand that, especially nowadays, concerns around data privacy and the misuse of AI are completely valid. That’s why we want to reassure you that only the research team will have access to this recording, and all audio files will be deleted once our analysis is complete, do you agree with this?

11.05 Survey Forms

User Experience Questionnaire (UEQ)

	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

Museum Experience Scale (MES) (Brochure)

Engagement	Knowledge/Learning
I enjoyed visiting the exhibition	The information provided by the brochure about the exhibits was clear
I felt engaged with the exhibition	I could make sense of most of the things and saw and did at the exhibition
The brochure helped me to feel mentally or physically engaged with the artworks or displays.	The brochure helped me to learn new things during the exhibition.
I felt I was experiencing the exhibition, rather than just visiting it	The brochure enriched my knowledge and understanding about specific exhibits
The brochure helped me to interact meaningfully with the content.	I discovered new information from the exhibits
Meaningful Experience	**Emotional Connection**
The information provided by the brochure made me feel that I was able to reflect on the significance of the exhibits and their meaning	The exhibition enabled me to reminisce about my past
During my visit, I put a lot of effort into thinking about the exhibition	The brochure helped me to interact meaningfully with the content.*
Seeing rare exhibits gave me a sense of wonder about the exhibition	The brochure helped me to feel emotionally connected to the artworks or exhibits.*
After reading the brochure, I was still interested to know more about the topic of the exhibition	I wanted to own exhibits like those that I saw in the exhibition
Seeing real exhibits of importance was most satisfying aspect of my visit to the exhibition	I felt connected with the exhibits

The items were presented in the scales as Likert items from 1 meaning *"strongly disagree"* to 7 meaning *"strongly agree"*.

Museum Experience Scale (MES) Chatbot

Engagement	**Knowledge/Learning**
I enjoyed visiting the exhibition	The information provided by the Chatbot about the exhibits was clear
I felt engaged with the exhibition	I could make sense of most of the things and saw and did at the exhibition
The Chatbot helped me to feel mentally or physically engaged with the artworks or displays.	The Chatbot helped me to learn new things during the exhibition.
I felt I was experiencing the exhibition, rather than just visiting it	The Chatbot enriched my knowledge and understanding about specific exhibits
The Chatbot helped me to interact meaningfully with the content.	I discovered new information from the exhibits
Meaningful Experience	**Emotional Connection**
The information provided by the Chatbot made me feel that I was able to reflect on the significance of the exhibits and their meaning	The exhibition enabled me to reminisce about my past
During my visit, I put a lot of effort into thinking about the exhibition	The Chatbot helped me to interact meaningfully with the content.*
Seeing rare exhibits gave me a sense of wonder about the exhibition	The Chatbot helped me to feel emotionally connected to the artworks or exhibits.*
After reading the brochure, I was still interested to know more about the topic of the exhibition	I wanted to own exhibits like those that I saw in the exhibition
Seeing real exhibits of importance was most satisfying aspect of my visit to the exhibition	I felt connected with the exhibits

The items were presented in the scales as Likert items from 1 meaning *"strongly disagree"* to 7 meaning *"strongly agree"*.

11.06 Participant Consent Form

Purpose of the Study

You are invited to participate in a research study examining how different methods of presenting information can influence visitor experiences in art exhibitions. Specifically, we aim to explore how digital tools, such as chatbots, may shape your engagement with artworks.

What Participation Involves

If you agree to take part in this study, you will:

- View paintings by various artists.
- In one section, receive information via a digital assistant (chatbot).
- In another section, explore the artworks without any digital guidance.
- Complete a brief questionnaire (approximately 10–15 minutes).
- Participate in a short interview about your experience with the exhibition.

Estimated Total Time Commitment: 20–30 minutes.

Confidentiality

All of your responses will remain completely anonymous and will be used solely for academic research purposes. We will not collect any personal or identifying information.

For transcription purposes only, the interview will be audio-recorded. The recording will be securely stored, used exclusively for generating the transcript, and permanently deleted once the analysis of your responses is complete.

Voluntary Participation

Your participation is entirely voluntary. You are free to withdraw from the study at any time, without any penalty or consequences.

Risks and Benefits

There are no known risks associated with participating in this study. While you may not receive direct benefits, your insights will contribute to the improvement of exhibition design and the development of educational tools in museum contexts.

Consent Statement

By clicking *"I agree that my answers may be analyzed for this research,"* you acknowledge that you have read and understood the information above and voluntarily agree to participate.

11.07 Thematic Analysis

Thematic Analysis-Figma File

11.08 UEQ Results

UEQ Results ArtZcape.pdf

11.09 Work Distribution

Team member	Matriculation Number	Contribution
Kübra Aslan	**36237**	Preliminary research about the artist and artworks Introduction Background Research Objectives Scope and Research Questions Hypothesis Discussion Interpretation of Findings Design Implications for Museum Chatbots Limitations and Challenges All design and formatting work
Ümmü Gülsüm Ergin	**36270**	Background research of chatbots Chatbot design and architecture System development and prompt engineering System updates after pre-test Technical documentation Future improvements and scalability Conclusion and summary
Itzel G. Vizcaíno García	**35481**	Digitization of instruments and interview design. Methodology Data Analysis Thematic Analysis Results