Article

Designing Reenacted Chatbots to Enhance Museum Experience

Yeo-Gyeong Noh [1] and **Jin-Hyuk Hong** [1,2,*]

1   School of Integrated Technology, Gwangju Institute of Science and Technology, Gwangju 61005, Korea; ygnoh0210@gist.ac.kr
2   Artificial Intelligence Graduate School, Gwangju Institute of Science and Technology, Gwangju 61005, Korea
*   Correspondence: jh7.hong@gist.ac.kr

Abstract: The increased availability of chatbots has drawn attention and interest to the study of what answers they provide and how they provide them. Chatbots have become a common sight in museums but are limited to answering only simple and basic questions. Based on the observed potential of chatbots for history education in museums, we investigate how chatbots impact history education and improve the overall experience according to their appearance and language style. For this, we built three models, designed by factors on embodiment and reflection, and 60 sets of answer–questions, designed for the National Museum of Korea. We conducted a study with a total of 34 participants and carried out a variety of analyses covering individual learning styles, museum experience scales, gaze data, in-depth interviews and observations from researchers. We present various results and lessons regarding the effect of embodiment and reflection on the museum experience. Our findings show how people with different learning styles connect with chatbot models and how visitors' behavior in the museum changes depending on the chatbot model. Specifically, the chatbot model equipped with embodiment and reflection shows its superiority in enhancing the museum experience, in general.

Keywords: chatbot; embodiment; historical reenactment; learning style; museum experience

1. Introduction

From theorists to researchers, those who engage in history education seek to answer why history is taught and what utility is expected [1]. History education helps people develop insight into society by expanding our human self-understanding beyond time and space [2]. This development process requires an understanding of logical criticism, analytical skills, empathy, intuition, insight, and imaginative reasoning [3] and is directly related to how to cultivate the "creativity" that is in the spotlight in the educational field [4]. According to Cooper, historical exploration, known as a core method in history education, demands that learners infer and use the concepts of time, cause, result, and motivation to explain the past, and this process itself is creative [5].

In various places, such as schools and museums, history education actively employs reenactment, one of the historical exploration approaches, to help people receive historical facts through their own experience [6]. Through these reenactments, we can achieve one of the goals of history education: to infer and empathize with the emotional state of historical figures. Simulation games, role plays, historical writing, and history journals and drawings are the common methods of reenactment in practice [7–10]. Passive listening to lectures in the classroom is less effective than history exploration through reenactment, and history education experts often point out that museums are the best place to implement and express such creativity [3].

Education in museums has drawn more attention than common lectures, for museums are places where people can feel the achievement and joy of self-expression and engage in the composition of active and self-directed knowledge through various "self-directed"

learning activities [11]. Furthermore, while only a few people actively study history with books after finishing school, many continue to visit museums throughout their lifetime.

Various studies have been conducted on how to design the museum experience to enhance educational effectiveness and engagement. In particular, it has become important to develop strategies that can encompass individuals' diverse motivations, social contexts, and situations, which may affect the visitors' museum experience and behaviors [12]. There are numerous related studies, for example, on the museum experiences of a certain age group [13], survey methods to evaluate museum user experiences [14], or user behavior patterns according to circumstances [15].

Chatbots have been widely employed for users to understand user-centric interactions through situational awareness, personalization, and adaptation [16] and can elicit a user's self-disclosure through nonverbal dialogue behavior, social dialogue, and social bond formation [17–19]. As such, their potential to be utilized in history education is highlighted as a technology that may be successfully applied in museums. Chatbots can represent a persona with the concept of anthropomorphism, reflection or reenactment, much like reincarnating a person from the past. Through a simulation of a conversation with important people from the past, visitors can effectively associate themselves with the reasoning or the emotional state of their conversation partners. With chatbots, the degree of concentration or empathy can be measured differently by introducing embodiment, e.g., a trait that can make a chatbot resemble a machine or a human being. The process of deducing the emotions of characters in history changes to inference through dialogue, and users can have different experiences depending on the language style of the chatbot.

Despite these potentials, chatbot studies have mainly focused on the performance of understanding user queries and providing answers. The design of the overall representation of a chatbot as an interactive entity is undervalued. As stated before, it is very important to infer and sympathize with the emotional state of historical figures through reenactment in history education; museums are also working to enhance educational effectiveness. Therefore, in this paper, we investigate ways to design chatbots to enhance the effectiveness of history education with respect to the museum experience.

2. Background

2.1. Chatbot

Conversational agents, commonly known as chatbots, are software applications designed to communicate with humans through voice or text to perform certain tasks [20]. As the chatbots' design and function have been improved, they are currently applied for education and entertainment purposes. Ruan et al. [21] designed a Quizbot so that students could learn factual knowledge in science, safety, and English vocabulary in a more effective way by interacting with chatbots than by using traditional methods. Fabio et al. [22] investigated how to design chatbot architecture, manage conversations, and provide correct answers as part of an e-learning system. To enhance the educational effect, Fadhil et al. [23] designed an educational chatbot game called CiboPoli to teach children about a healthy lifestyle. With this kind of interactive social game environment, users can become more deeply engaged and immersed in learning [24,25].

In addition to developing chatbot systems that correctly answer users' input queries, some recent studies consider the personalization and appearance design of chatbots. A conversational system was created to automatically generate responses to user requests on social media by integrating deep learning techniques to show empathy when interacting with users [26]. An agent, Tinker, was designed to enhance engagement and learning effects in museums, appearing in the form of a human-sized anthropomorphic robot using nonverbal dialogue behaviors, empathy, social dialogues, and reciprocal self-disclosure to establish social bonds with its users [27]. Researchers have also discussed how to represent the conversational agent on the interface and how to design the interaction [28]. As a part of an embodied conversational agent (ECA), chatbots often feature an embodiment, giving the agent a human appearance [29]. A more human-like design may be accomplished

through a combination of facial expressions, speech, gestures, and body posture for more effective interaction [30,31].

2.2. Museum Experience

With the introduction of new technologies in museums, researchers have explored human behavioral patterns and how these technologies may improve visitors' experiences. To have a better understanding of how people enjoy and experience museums, many studies have investigated methods to measure visitors' subjective experience and to identify characteristics for specific situations or populations. Some researchers presented a preliminary framework for a nomadic web content design and evaluated handheld devices and user interaction to enhance the museum experience [32]. Significant references in the analysis included users' sense of isolation and users' attempts to seamlessly switch between reality and hypothetical contexts.

Technological advances have also led to a better museum experience, e.g., playing augmented reality/virtual reality games related to exhibitions that provide proper information and stimulate imagination and interest by making the exhibits livelier and more realistic. Based on an in-depth survey of users, researchers have also made suggestions related to costs, usability, and the quality of the sensorial experience by classifying virtual reality installations used in museums in terms of interaction and immersion [33]. For a specific user group, some researchers measured qualitative and quantitative data obtained during the use of two contents created through co-design with teenagers, which enabled them to map personality traits into different behaviors [34]. Analyses of user experiences related to mobile-based storytelling "guides" commonly used in museums have uncovered many issues in designing digital experiences for cultural, educational, and entertainment purposes. Studies have focused on themes that represent components, such as interactive story plots and narrative, staging and way-finding in the physical space, personalization, and social interaction [35–37].

2.3. Chatbot in GLAMs (Galleries, Libraries, Archives, Museums)

From the early "speaking machine" to the chatbot powered by artificial intelligence (AI), there have been numerous attempts to engage audiences in GLAMs [38]. Machidon et al. [36] developed a more accessible and easier way for the general public to search for information in large digital libraries by using a conversational agent. In addition, question and answers, frequently asked questions (FAQs), in tree structure can be archived by allowing chatbots to learn on their own [39,40]. Since chatbots interact naturally with people in natural language, they have the potential and capability to provide information and services on digital cultural heritage resources, similar to an experienced human docent [41,42]. For instance, the poignant museum in Amsterdam offers visitors a wider historical experience to discuss the Holocaust more fully by providing visitors with the opportunity to see Anne Frank and her family's home during their years of hiding from Nazi persecution [43]. The Field Museum in Chicago, IL, USA., led the participation by introducing chatbot technology to allow visitors to talk to the dinosaur Máximo [44]. The Carnegie Museum in Pittsburgh has developed a gamified chatbot called Andy Carnegie Bot, which allows visitors to collect digital stamps using Facebook messenger [45].

Researchers have reported three important research topics that can be explored to develop practical technology solutions for using chatbots in museum environments: information processing, multimodal intent detection, and dialogue management [46]. As AI chatbots are a key technology for implementation in museums, recent studies have examined approaches and ways to provide solutions for natural language translation, knowledge bases, and linked open data source inquiry [37]. A conversational agent also has advantages in providing interesting strategies to engage young audiences to first-person narratives [47]. In addition, studies on museum experience design have noted the possibility of chatbots using storytelling techniques rather than simply being used as messaging apps [48]. By delivering information on cultural heritage and artifacts through the

storytelling method, a museum can effectively display artifacts and, as a result, attract and expand various audiences. For example, storytelling that evokes emotions in the audience can be achieved by introducing not only the work of an artist, but also the story of his/her life, the historical and environmental context in which they lived, and related objects.

3. Research Questions

Museums currently exert much effort to engage visitors, deliver knowledge, create meaningful experiences, and make emotional connections. Since the museum experience is influenced not only by the museum itself, but also by individual motivation and attitudes, proper interaction and connection between the museum and visitors is required to maximize their museum experience. Based on the possibility of using chatbots in the field of museums and history education, in this study, we investigated a way to connect visitors and chatbots that provides a service supporting history education in museums. For this, we designed three chatbot models considering the core functions of embodiment and reenactment and evaluated them with various measurements, including UEQ (User Experience Questionnaire) and MES (Museum Experience Scale). With these motivations and problem statements, we set up the following two research questions:

- RQ1: What factors of the chatbot design influence the users' interaction and museum experience?
- RQ2: What are the specific relationships between individuals' learning style and museum experience?

4. Implementation

We constructed a knowledge base on 'Silla (from 57 AD to 935 BC)', one of the three kingdoms in ancient Korea, to organize chatbot conversations into topics that are important to the National Museum of Korea. The Kingdom of Silla is renowned for its wealth of historical monuments and relics resulting from its achievement of conquering the kingdoms of Baekjae and Goguryeo, thus unifying the Korean Peninsula; a number of historical studies have highlighted the importance of Silla-related content in Korean history education [13,49]. We divided the contents into general knowledge of Silla and specific knowledge of King Jinheung. To carefully convey this knowledge, we considered two factors when designing the chatbots: the content discussed and the model with which the user speaks.

4.1. Content Design

To determine what should constitute the knowledge base, we extracted frequently asked questions in search engines by web crawling and collected a number of questions and answers about relics, culture, and historical figures related to Silla. We have comprehensively included cultural and political background information related to Silla, which could represent common topics in Korean history education, as well as information related to a particular person, to demonstrate its effectiveness. The results indicated that King Jinheung, the 24th monarch of Silla, is the most frequently queried historical figure; thus, we composed specific knowledge about facts and stories related to him. A total of 259 pairs of questions and answers were collected, of which 147 pairs and 112 pairs were constructed on Silla and King Jinheung, respectively. As shown in Table 1, each type of knowledge comprises three categories. In detail, the content about general knowledge covers the following subjects: (1) people who existed in the Silla kingdom; (2) the lifestyle of the times, such as food, clothing, and shelter; and (3) relics from this time period, such as crowns, monuments, and ceramics. The following content comprises specific knowledge about King Jinheung: (1) personal information about the king, e.g., the meaning of his name and his age at death; (2) historical events that the king was involved in; and (3) the political system implemented by King Jinheung.

Table 1. Historical content.

General Knowledge (Kingdom of Silla)	Specific Knowledge (King Jinheung)
G1. Historical figures	S1. Information about the king
G2. Lifestyles	S2. Historical events
G3. Relics	S3. Politics and systems

4.2. Model Design

To explore the research questions, as shown in Figure 1 and Table 2, we designed three chatbot models in consideration of three factors: embodiment (humanoid), reenactment, and language style. In practice, two common methods of information acquisition are utilized in museums: searching for information through kiosks and listening to explanations given by a docent. The question–answering (QA) system mirrors the kiosk-based interaction, and the docent imitates a human docent in museums. In addition to the two common models, we designed a historical figure (king) that returns from the past to inform users about historical content. All contents, question–answer pairs for chatbots, and the design of models were examined for authenticity by a curator in the National Museum of Korea. The two embodied chatbots, the docent and the historical figure, were created in 3D and have five motions (shaking hands, bowing, pointing, arm-twisting, and agonizing) to reflect the embodiment characteristics. The answers of the three models were provided to the users by both text and audio. The answer audio was generated by using oddcast (https://ttsdemo.com/, accessed on 7 May 2021), a text-to-speech program. The historical figure has an additional five facial expressions (joy, surprise, anger, sadness, and disgust) to express his emotions depending on the answer.

Figure 1. Appearance of three chatbot models.

Table 2. Three chatbot models designed.

Chatbot Model	Embodiment	Reenactment	Language Style	
QA system	Not Embodied	Not Reenacted	Third-person point of view	Phrase
Docent	Embodied	Not Reenacted	Third-person point of view	Modern sentence
Historical figures	Embodied	Reenacted	First-person point of view	Archaic sentence

4.2.1. QA System

Similar to the kiosk search engine, the QA system adopts neither embodiment nor reflection. The QA system adopts a concise phrase form to deliver information, and it has a third person's perspective, allowing it to answer objectively, e.g., "King Jinheung's representative achievements: expanding territory, encouraging Buddhism, and promoting cultural industries. Sending Isabu and Sadaham to destroy Daegaya and establish Silla's dominion."

4.2.2. Docent

Adopting the characteristics of the commentator in museums, the docent, assuming human form, speaks in full sentences and explains the factual information objectively from an observer's viewpoint. As if she lived in modern times, her answer is made up of a modern sentence, e.g., "King Jinheung expanded Silla territory as a policy of conquering, encouraged Buddhism, and promoted the cultural industry. In addition, he sent Isabu and Sadaham to destroy Daegaya and establish Silla's dominion."

4.2.3. Historical Figure

The most distinctive feature of the historical figure is reenactment, a reflection of the past. As a person who comes back from the past and talks to people in modern-day museums, the historical figure adopts the appearance common in his lifetime, e.g., hairstyle and clothing. He also speaks in the historical dialect, using archaic words, e.g., "I expanded mine (my) eard's (country's) territory widely through mine (my) excellent conquest policy. I encouraged Buddhism and promoted various cultural industries. Especially, I am most (very) proud of Isabu and Sadaham who abreotan (destroyed) Daegaya and established Silla's dominion for me" (note that this example is translated similarly to old English). The garment he wears also reflects the culture of the time according to historical evidence, and only he adopts the first-person viewpoint, relaying the historical information as if he experienced it in person.

5. Methodology

5.1. Participants' Demographics

The participants in our study included 19 men and 15 women all in their 20s, with an average age of 24.1 years. Due to coronavirus disease 2019 (COVID-19), we recruited participants as insiders because we had to conduct experiments according to the institute's policy. The participants were evenly distributed in terms of their interest and understanding of history. In a preliminary survey on history education, we confirmed that the average interest in history was 3.66 points on a scale of 1–7, and the background knowledge about history was 2.48 points on a scale of 1–5. According to the preliminary survey on the museum, common behaviors of visitors included taking pictures of exhibits or using display tools, such as kiosks. In detail, museum users reported the following: taking a picture of an exhibit, 34.6%; using the display tools provided in kiosks and other multimedia, 30.7%; renting exhibition tools, such as voice commentary, 15.3%; listening to a docent's description, 11.5%; and taking notes on the exhibition, 7.6%. Therefore, the experiment was conducted with those who have evenly distributed interest in history and education and are familiar with using display or exhibition tools in museums.

5.2. Measures

5.2.1. Kolb's Learning Style

Learning styles refer to ways that individuals consistently prefer to collect and process information. Kolb proposed 'experiential learning theory'—the idea that knowledge stems from the perception and processing of experience [50]. The way in which a person perceives information can be divided into concrete experience (CE: pragmatist) or abstract conceptualization (AC: theorist), while how a person processes information can be classified as active experimentation (AE: activist) or reflective observation (RO: reflector). Based on this, many education experts identify students' learning traits and devise teaching methods accordingly. Pragmatists tend to prefer explanations through specific examples or learning to discuss and present with colleagues. Theorists tend to be analytical, logical and oriented towards symbols. In addition, they learn most effectively in authority- directed, impersonal learning situations that emphasize theory and systematic analysis. Activists have extravert tendencies, so they rely heavily on experimentation and learn best while engaging in projects. Last, reflectors tend to be introverts who depend on careful observation when making judgments.

5.2.2. User Experience Questionnaire (UEQ)

The UEQ aims to measure a comprehensive impression of the user experience [51]; we use this to evaluate the chatbot interaction in general. The survey evaluates six factors: 'attractiveness' on overall impressions, 'perspicuity' on whether users can easily learn how to interact with the chatbot, 'efficiency' on whether the desired task can be performed quickly, 'dependability' on whether users feel they can control the interaction, 'stimulation' on whether they feel interested or motivated to use the item, and 'novelty' on whether the interaction is creative or not.

5.2.3. Museum Experience Scale (MES)

The MES is a museum-specific survey designed to examine the impact of multimedia guides on the user experience in cultural spaces [52]. It consists of four main components— engagement, knowledge/learning, meaningful experience, and emotional connection— each of which consists of five questions.

5.3. Procedure

The experiment was originally scheduled to take place at the National Museum of Korea, but it was conducted in a separate space in a laboratory because the museum was closed due to the coronavirus disease 2019 (COVID-19) pandemic. To help users feel like they were in a museum, our application showed many pictures taken inside the museum. The experiment was conducted as a within-group design, and the three chatbot models were presented to the participants in a counterbalanced order. From 42 to 60 question–answer sets were presented to each participant through six combinations of three chatbot models and two content types. By mixing the order of model and content, we minimized the impact of the order of the combinations.

Through the presurvey, we collected information on the demographics and learning styles of the participants. During the experiment, the participants were asked to complete the UEQ and MES surveys every time they completed a set of conversations with a chatbot model. After completing all the surveys on the three chatbots, they were additionally asked to choose which chatbot they preferred most for each of the six contents. The conversation with each chatbot took approximately 8–10 min, and it took another 5 min to complete the survey. Participants spent 40 min, on average, completing the whole process and received USD 15 for their time and contribution. In addition, qualitative data from interviews and researchers' observations and participants' gaze data from the gaze tracker (Tobii Pro Nano) were collected to analyze their behaviors and in-depth responses.

6. Analysis and Results

6.1. Chatbot Interaction: UEQ

The UEQ was used to evaluate the chatbot interaction. Three pairs of the three models were analyzed through paired *t*-tests. Among the six factors evaluated by the UEQ, attractiveness, stimulation and novelty had significant results depending on the model, but no difference was found in perspicuity, efficiency and dependability.

As shown in Figure 2, the QA and the king showed significant differences ($p < 0.001$) in terms of attractiveness, stimulation, and novelty. When comparing the QA and the docent, there were also significant differences ($p < 0.01$) in the three factors. In the comparison between the king and the docent, there was no statistically significant difference for all factors. This statistical analysis showed significant differences in the king—QA and docent— QA analyses, which can be interpreted as a result of differences in the application of embodiment. For the other three factors, we found no difference between models, but perspicuity, followed by dependability and efficiency, had the highest scores; all of these points were higher than neutral. This means that the interaction enables the users to easily learn how to use the system, makes users feel comfortable because they can control the system and is characterized by fast performance of the desired task. With respect to novelty, only the king presented a positive experience.
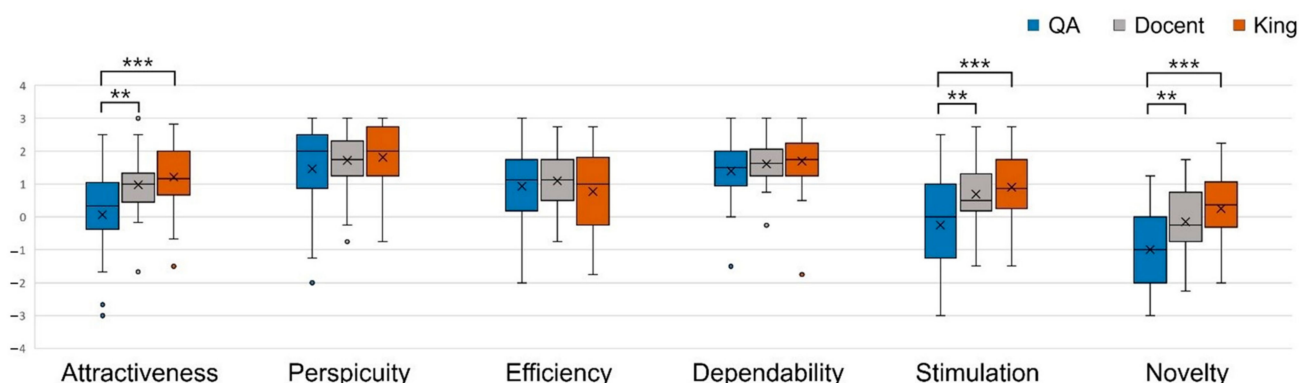
Figure 2. UEQ results on the three chatbot models. (significance *p*-value **: $p < 0.01$, ***: $p < 0.001$).

6.2. Museum Experience: MES

6.2.1. Chatbot

Engagement:

As shown in Figure 3, there were significant differences in all relationships between the three models, with the king scoring the highest, followed by the docent and the QA ($p < 0.001$). In other words, when a chatbot is used to listen to the explanation related to the exhibition, the degree of engagement clearly varies depending on the chatbot model. This variation has a comprehensive impact on the elements set up based on the design of the chatbots. For the presentation, one participant said in an interview, (P12) "Since it was an incident that King Jinheung experienced in person, I was able to concentrate better because it seemed to tell a story."

Knowledge/Learning:

There was a significant difference when comparing the king with the docent ($p < 0.001$) and the king with the QA ($p < 0.001$). This means that the docent and QA received similarly high scores, while the king scored relatively low. Many participants stated in post interviews that the docent was more professional and reliable when delivering knowledge. Regarding the QA model, one participant also highlighted the advantages of the QA in conveying knowledge, stating, (P25) "I liked the QA model better than the other models because it was simpler and easier to read than the long and detailed description of the other two models." In general, the docent had a higher knowledge/learning score. This can be supported by the study of embodiment, which has shown that people tend to feel that a source is more reliable because of its elements [31].

Meaningful Experience:

This factor is an indicator of the meaningful experience created by appreciating the exhibition. There was no significant difference in this factor, which was expected to be due to limitations in the experimental environment. A meaningful experience is often dependent on interpersonal differences, such as interest in content or background knowledge. However, in this experiment, the model and contents were counterbalanced to reduce the impact on content.

Emotional Connection:

This factor had the same score and significance ($p < 0.001$) as the engagement model. One participant (P5) emphasized the importance of historical reenactment, saying, "When learning about the motivation or decision-making process for events, I have a richer understanding and empathy of events when they are explained by King Jinheung rather than by the docent."

As a result of analyzing how the museum experience varies depending on the chatbot model, engagement and emotional connection, two of the MES factors, were found to be significant in the order of the king, the docent, and the QA. In terms of knowledge/learning,

the ranking was reversed to the docent, the QA, and the king. To obtain a closer look at how some of the characteristics of the chatbot model affect users, we divided the users by type of learning style and analyzed the results.
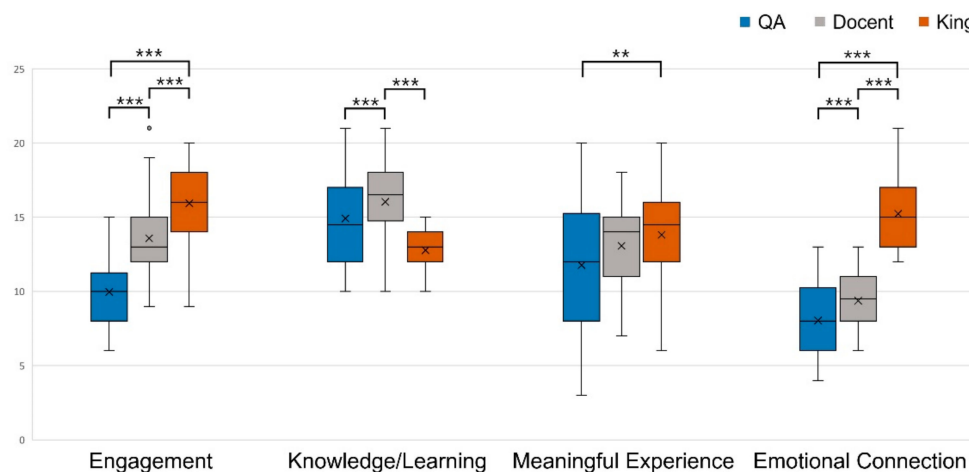

Figure 3. MES results on the three chatbot models. (significance *p*-value **: $p < 0.01$, ***: $p < 0.001$).

6.2.2. Chatbot and Learning Style

We analyzed the museum experience by dividing the participants by the type of Kolb learning style, as shown in Figure 4. As a result, we found that using the same model provides different stimuli and experiences.

Activist vs. Reflector:

The analysis of users divided into activists and reflectors showed significant results in terms of engagement and emotional connection. In both factors, the QA received the lowest score, and the king received the highest score. The engagement scores were significantly different among all three models, with activists scoring higher than reflectors: the QA ($p < 0.001$), the docent ($p < 0.01$), and the king ($p < 0.001$). However, in terms of emotional connection, reflectors scored higher than activists, and among the three models, only the QA ($p < 0.001$) had significant results.

Pragmatist vs. Theorist:

The MES was analyzed for pragmatists and theorists according to how information was processed, and a significant difference was identified in terms of knowledge/learning. The docent received the highest score, followed by the QA and the king, and theorists had higher scores than pragmatists in all cases. Only the QA ($p < 0.001$) and the king ($p < 0.05$) were validated for significance in user-specific analysis, especially in the case of using the QA, which noticeably increased theorists' knowledge scores.

According to the two aspects of learning styles, individuals were sorted into four groups: 10 activist and pragmatist, 9 activist and theorist, 8 reflector and pragmatist, and 7 reflector and theorist individuals. As shown in Figure 5, Group 1 (which represents the intersection of activist and pragmatist) and Group 2 (which represents the intersection of activist and theorist) had extremely high preferences for the king among the four groups because they were mainly influenced by the aspects of engagement and emotional connection, as discussed earlier. The group with the highest preference score for the docent was Group 4 (which represents the intersection of reflector and theorist), with an emphasis on knowledge/learning. However, given the king's MES results, Group 3 (which represents the intersection of reflector and pragmatist), which was expected to have higher preference scores for the king, unexpectedly preferred the QA and the docent. The reason for this result is that the chatbot preferences cannot be absolutely determined by the learning style, and it is highly likely that the four elements of the MES do not have the same weight independently.
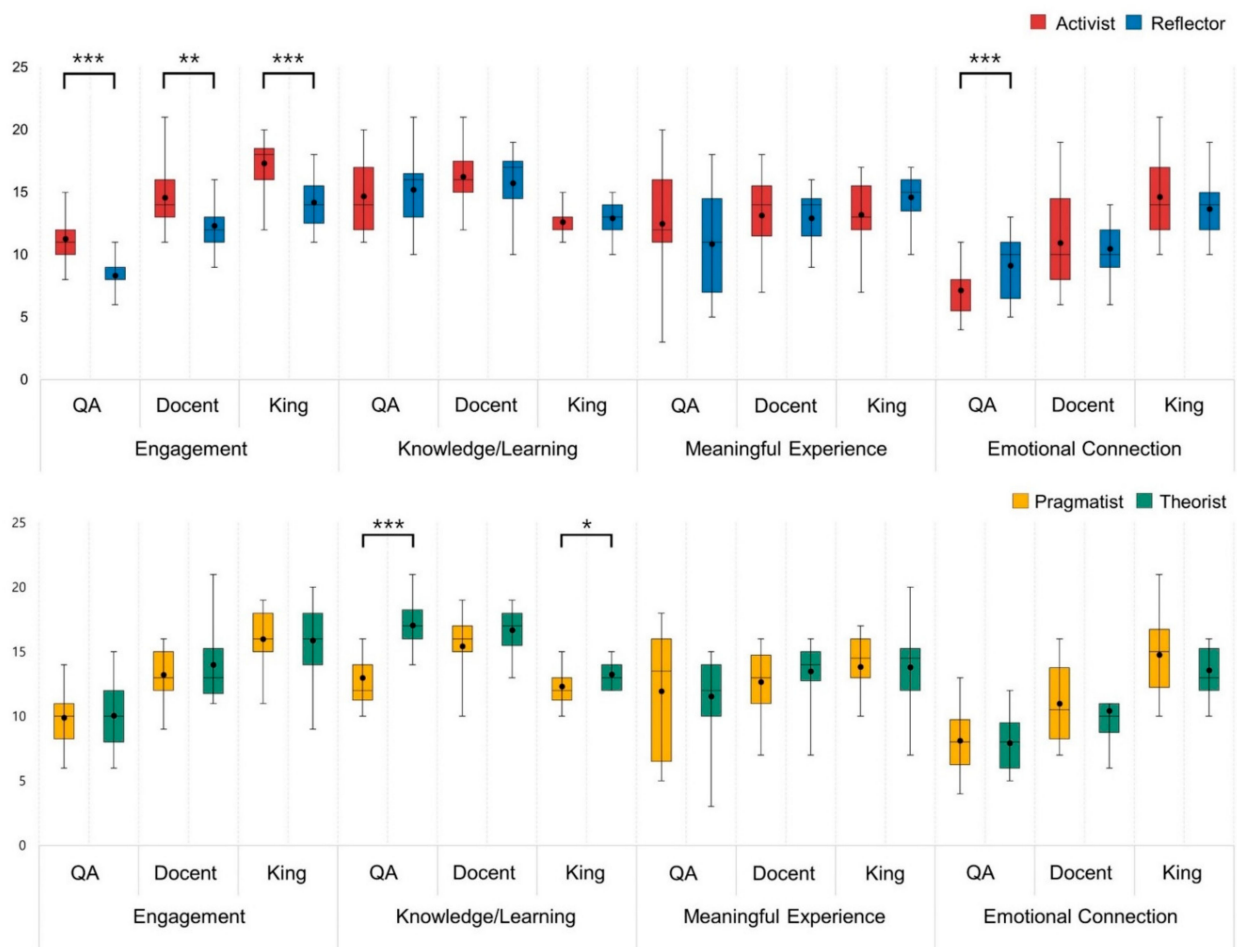
Figure 4. MES results with respect to learning style. (upper: Activist vs. Reflector, lower: Pragmatist vs. Theorist). (significance *p*-value *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).
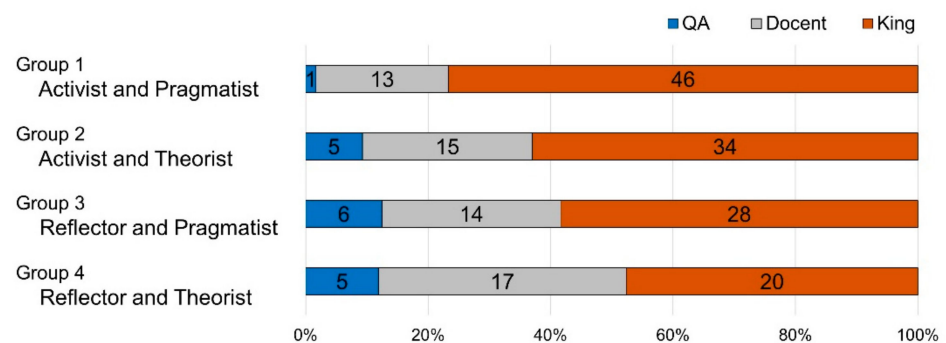

Figure 5. Best model preference by learning style.

Analysis with Gaze and Behavior Data:

Eye tracking data were collected during all experiments using Tobii Pro Lab software. Among the basic units of gaze data (fixation count, fixation duration, saccades), we conducted an analysis with a fixation count to assess which AOI (area of interest) received more or less attention. The AOI consists of content topic, model, selected question, answer to question, and the next questions users can choose from: theme, model, question, answer, and options.

We analyzed two aspects of the visual information: where the user's focus is concentrated according to the model and which behavioral characteristics appear by learning style. The QA model tended to draw more attention to the 'answer' to questions and the

following questions 'options' than other AOIs. This can be interpreted in conjunction with the aforementioned MES results. As presented in Figure 6, users who used the QA, which emphasized the knowledge/learning aspect, tended to focus on 'answer'. Moreover, it is assumed that the reason the AOI for the following questions (options) had a high fixation count value is that the answers were relatively concise, so it took time to choose the next question after viewing the answers. For the docent model, all AOIs had intermediate values among the three models. When using the docent model, the user looked at all the AOIs evenly without bias. When interacting with King Jinheung, the level of attention focused on the model was significantly higher, which may be because the King Jinheung model had the highest score of engagement and emotional connection.
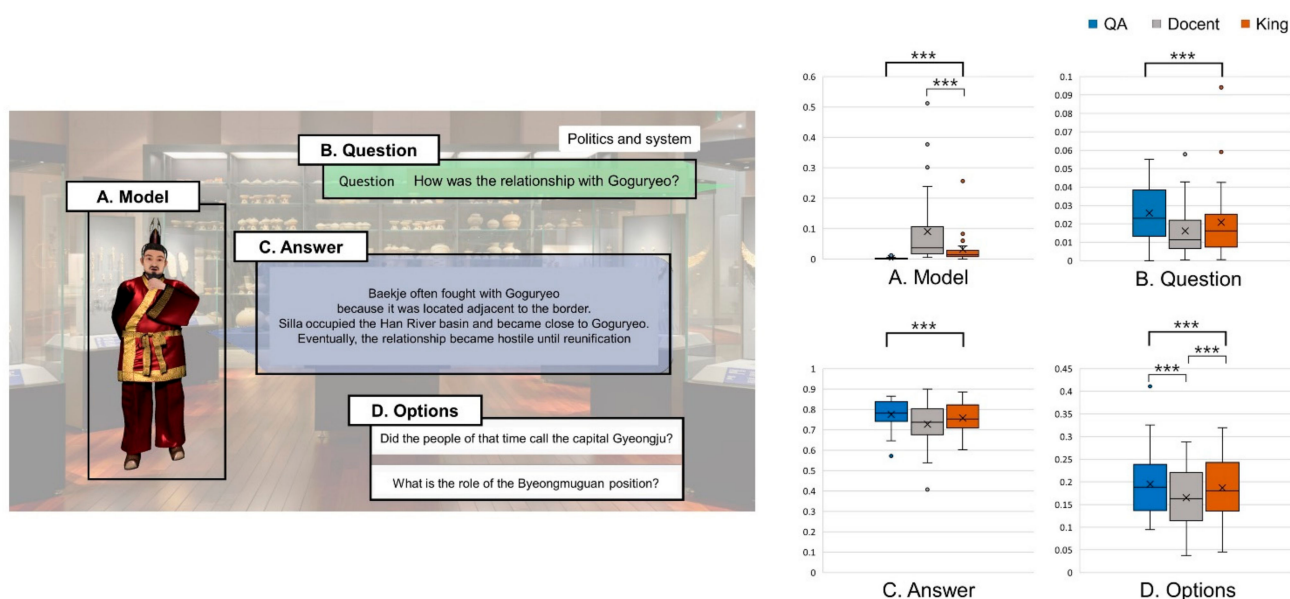

Figure 6. Area of interest for gaze tracking and gaze results on the three chatbot models. (significance *p*-value ***: $p < 0.001$).

In regard to learning style, activists paid more attention to the model than reflectors ($p < 0.001$), while reflectors paid more attention to the theme ($p < 0.01$). This may result from activists' extroverted inclination or reflectors' propensity to identify the topic they are talking about. Interestingly, theorists focused on 'question' and 'answer', which, due to their analytical and logical characteristics, suggests that they are more attracted to the knowledge itself than to the interaction with any chatbot model.

7. Discussion

7.1. What User Type Best Matches the Chatbot Model?

We designed the content to be generally user-friendly, with topics of interest and frequently asked questions. For this study, we tried to minimize the impact of content to better analyze the impact of factors such as the embodiment, reflection, and language style of the chatbot models on the museum experience. However, through a chatbot preference survey and qualitative interview data, we found a significant relationship between the content and the chatbot models. S1, S2 and S3 were all grouped into the same category with content for a specific person, but each had a slightly different chatbot preference. In particular, as shown in Figure 7, in the case of S2 and S3, the king was consistently preferred, which may be because the content itself matches the chatbot regardless of the learning style. In the case of G1, the preference for the chatbot model was heavily influenced by the learning style of participants. In particular, we found that all participants in Group 2 preferred the king, while most participants in Group 4 preferred the QA or the docent. When recommending a chatbot model for a user, it is necessary to consider both content and learning style.
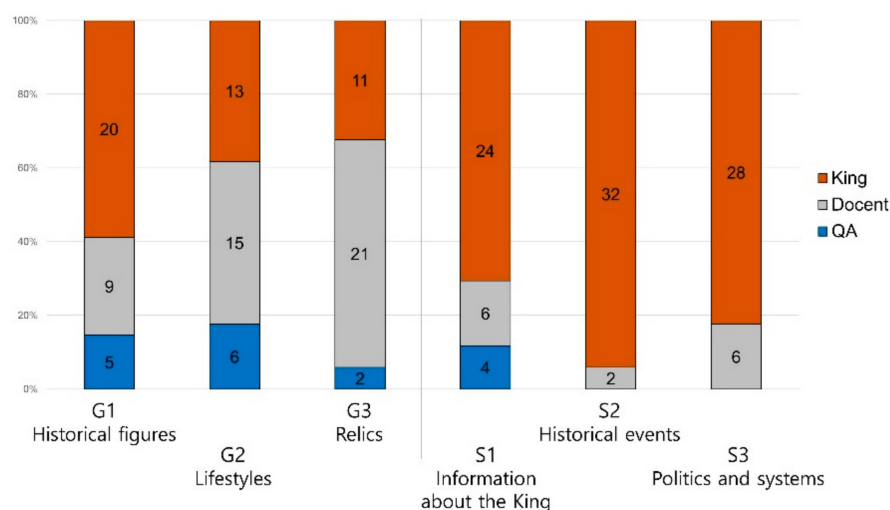
Figure 7. Model preference on content.

7.2. User Behaviors while Interacting with the Chatbots

The analysis of museum behavior with gaze data and the observation of the researcher showed that the participants were represented by three types in relation to the learning style. The first type is the 'responsive communicator', who nods or tilts his/her head, shouts 'uh-huh' or laughs at funny answers while listening to the chatbot's response. This type often appeared in people with an activist learning style. An evaluation of the gaze data further indicated a tendency to look at the avatar of the chatbots or read the answers with the eyes, without diverging attention elsewhere. The next type is the 'quick receiver', who glances at the answer and chooses the next question before the audio finishes. This type of user moves on to another chatbot with a few turns of conversation. The last type is the 'enthusiastic seeker'. This type looks for other visual aids related to the conversation or clicks on models, buttons, and backgrounds to see if there are any other functions. They also ask many additional questions about the technology or functions used to create the program.

7.3. Concerns Regarding the Bias of Presenting a Historical Figure

Historical events are stories intertwined with the interests of many countries and the personal opinions of a wide range of individuals. By making a historical figure into a chatbot, the emotional connectivity improved in the user's museum experience, and the interviews showed that for some users, the chatbot enabled the user to empathize with a particular person's behavior. As mentioned above, supporting empathy in history education may develop an individual's ability to understand an unfamiliar past that is unknown to the learner by improving creativity. However, there are also concerns about the bias of historical thought, especially from education experts. This is because users can only empathize with the opinion of the people (chatbots) they have spoken with. The presentation of multiple chatbots of various figures could provide a diversity of opinions, or in situations where objectivity is suspect, we can protect users from historical distortions and bias by supplying an answer from a docent, not the historical figure.

8. Conclusions

In this study, we evaluated three chatbot models designed based on the concepts of embodiment and reenactment as historical information providers to enhance the museum experience. With 34 participants, we analyzed the three chatbots in terms of the UEQ, the MES, gaze data, and interviews, where we confirmed that embodiment and reenactment strongly influence the three components of the museum experience. Specifically, embodiment contributes solely to knowledge/learning, while engagement and emotional connection are highly attributed to both embodiment and reenactment. We also examined

whether learning style and content have an effect on the choice of a chatbot model, i.e., there is no perfect chatbot model that can be applied to every situation.

As mentioned above in the background section, the information provided by the museum should be objective, factual and validated by experts. Therefore, chatbots are often created using rule-based methods, which have limitations in generating flexible answers. Since this paper dealt with specific subjects of history education, it is still necessary to verify our results on a greater range of subjects in museums. Additionally, the experiments were conducted in a limited environment due to the COVID-19 pandemic, so it is necessary to verify this platform in an authentic museum environment with appropriate sites and relics outside of the laboratory. Furthermore, considering the possibilities of AR/VR, we will study the application of the 3D model and speech source used in the above experiments in various platforms in future work.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Levstik, L.S.; Barton, K.C. *Researching History Education: Theory, Method, and Context*, 1st ed.; Routledge: New York, NY, USA, 2018.
2. Levstik, L.S.; Barton, K.C. Committing acts of history: Mediated action, humanistic education and participatory democracy. In *Critical Issues in Social Studies Research for the 21st Century*; Stanley, W., Ed.; Information Age Publishing: Greenwich, CT, USA, 2001; pp. 119–147.
3. Marcus, A.S.; Stoddard, J.D. *Teaching History with Museums: Strategies for K-12 Social Studies*; Taylor & Francis: Abingdon, UK, 2017.
4. Wilson, A. *Creativity in Primary Education*, 3rd ed.; SAGE Publications: Thousand Oaks, CA, USA, 2009.
5. Jackson, N. *Creativity in History Teaching and Learning. Subject Perspectives on Creativity in Higher Education Working Paper*; The Higher Education Academy: Heslington, UK, 2005; Available online: https://www.creativeacademic.uk/uploads/1/3/5/4/13542890/creativity_in_history.pdf (accessed on 7 August 2021).
6. Agnew, V. History's affective turn: Historical reenactment and its work in the present. *Rethink. Hist.* **2007**, *11*, 299–312. [CrossRef]
7. Aidinopoulou, V.; Sampson, D. An action research study from implementing the flipped classroom model in primary school history teaching and learning. *J. Educ. Technol. Soc.* **2017**, *20*, 237–247.
8. Yilmaz, K. Historical empathy and its implications for classroom practices in schools. *Hist. Teach.* **2007**, *40*, 331–337. [CrossRef]
9. Bernik, A.; Bubaš, G.; Radošević, D. Measurement of the effects of e-learning courses gamification on motivation and satisfaction of students. In Proceedings of the 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2018), Opatija, Croatia, 21–25 May 2018; pp. 0806–0811. [CrossRef]
10. Sreelakshmi, A.S.; Abhinaya, S.B.; Nair, A.; Nirmala, S.J. A Question Answering and Quiz Generation Chatbot for Education. In Proceedings of the Grace Hopper Celebration India (GHCI 2019), Bangalore, India, 6–8 November 2019; pp. 1–6.
11. Hein, G.E. Museum education. In *A Companion to Museum Studies*; Macdonald, S., Ed.; Blackwell Publications: Malden, MA, USA, 2006; pp. 340–352.
12. Falk, J.H.; Dierking, L.D. *Learning from Museums: Visitor Experiences and the Making of Meaning*; AltaMira Press: Walnut Creek, CA, USA, 2000.
13. Chung, D.J. The effects of history education activities using Shilla founder's myth on children's historical attitude and creative leadership. *J. Res. Inst. Silla Cult.* **2018**, *52*, 1–23. [CrossRef]
14. Pallud, J.; Monod, E. User experience of museum technologies: The phenomenological scales. *Eur. J. Inf. Syst.* **2010**, *19*, 562–580. [CrossRef]
15. Zancanaro, M.; Kuflik, T.; Boger, Z.; Goren-Bar, D.; Goldwasser, D. Analyzing museum visitors' behavior patterns. In *User Modeling 2007*; Conati, C., McCoy, K., Paliouras, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 238–246.
16. Chaves, A.P.; Gerosa, M.A. How should my chatbot interact? A survey on human-chatbot interaction design. *Int. J. Hum. Comput. Interact.* **2019**, *37*, 729–758. [CrossRef]
17. Ho, A.; Hancock, J.; Miner, A.S. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *J. Commun.* **2018**, *68*, 712–733. [CrossRef] [PubMed]

18. Lee, Y.C.; Yamashita, N.; Huang, Y.; Fu, W. "I Hear You, I Feel You": Encouraging deep self-disclosure through a chatbot. In Proceeding of CHI Conference on Human Factors in Computing Systems (CHI 2020), Honolulu, HI, USA, 25–30 April 2020; pp. 1–12.

19. Van der Lee, C.; Croes, E.; de Wit, J.; Antheunis, M. Digital confessions: Exploring the role of chatbots in self-disclosure. In Proceeding of Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, 19–20 November 2019; p. 21.

20. AbuShawar, B.; Atwell, E. ALICE chatbot: Trials and outputs. *Comput. Sist.* **2015**, *19*, 625–632. [CrossRef]

21. Ruan, S.; Jiang, L.; Xu, J.; Tham, B.J.K.; Qiu, Z.; Zhu, Y.; Murnane, E.L.; Brunskill, E.; Landay, J.A. QuizBot: A dialogue-based adaptive learning system for factual knowledge. In Proceeding of the CHI Conference on Human Factors in Computing Systems (CHI 2019), Glasgow, Scotland, 4–9 May 2019; p. 357.

22. Wu, E.H.; Lin, C.; Ou, Y.; Liu, C.; Wang, W.; Chao, C. Advantages and Constraints of a Hybrid Model K-12 E-Learning Assistant Chatbot. *IEEE Access* **2020**, *8*, 77788–77801. [CrossRef]

23. Ondáš, S.; Pleva, M.; Hládek, D. How chatbots can be involved in the education process. In Proceeding of the 2019 17th International Conference on Emerging eLearning Technologies and Applications (ICETA), Starý Smokovec, Slovakia, 21–22 November 2019; pp. 575–580.

24. Clarizia, F.; Colace, F.; Lombardi, M.; Pascale, F.; Santaniello, D. Chatbot: An education support system for student. In *Cyberspace Safety and Security*; Castiglione, A., Pop, F., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 291–302.

25. Fadhil, A.; Villafiorita, A. An adaptive learning with gamification & conversational UIs: The rise of CiboPoliBot. In *25th Conference on User Modeling, Adaptation and Personalization (UMAP 17), Bratislava, Slovakia, July 2017*; Tkalcic, M., Thakker, D., Eds.; Association for Computing Machinery: New York, NY, USA, 2017; pp. 408–412.

26. Hu, T.; Xu, A.; Liu, Z.; You, Q.; Guo, Y.; Sinha, V.; Luo, J.; Akkiraju, R. Touch your heart: A tone-aware chatbot for customer care on social media. In *CHI Conference on Human Factors in Computing Systems (CHI 2018), Montreal QC, Canada, April 2018*; Mandryk, R., Hancock, M., Eds.; Association for Computing Machinery: New York, NY, USA, 2018; p. 415.

27. Bickmore, T.; Pfeifer, L.; Schulman, D. Relational agents improve engagement and learning in science museum visitors. In *Intelligent Virtual Agents*; Vilhjálmsson, H.H., Kopp, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 55–67.

28. McTear, M.; Callejas, Z.; Griol, D. *The Conversational Interface: Talking to Smart Devices*; Springer International Publishing: Cham, Switzerland, 2016.

29. Takeuchi, A.; Naito, T. Situated facial displays: Towards social interaction. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '95), Denver Colorado, USA, 7–11 May 1995*; Katz, I.R., Mack, R., Eds.; ACM Press: New York, NY, USA, 1995; pp. 450–455.

30. Becker, C.; Kopp, S.; Wachsmuth, I. Why emotions should be integrated into conversational agents. In *Conversational Informatics: An Engineering Approach*; Nishida, T., Ed.; John Wiley & Sons, Ltd.: Chichester, UK, 2007; pp. 49–68.

31. Cassell, J.; Bickmore, T.; Billinghurst, M.; Campbell, L.; Chang, K.; Vilhjálmsson, H.; Yan, H. Embodiment in conversational interfaces: Rea. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, Pittsburgh, PA, USA, 15–20 May 1999; pp. 520–527.

32. His, S. A study of user experiences mediated by nomadic web content in a museum. *J. Comput. Assist. Learn.* **2003**, *19*, 308–319.

33. Carrozzino, M.; Bergamasco, M. Beyond virtual museums: Experiencing immersive virtual reality in real museums. *J. Cult. Herit.* **2010**, *11*, 452–458. [CrossRef]

34. Cesário, V.; Petrelli, D.; Nisi, V. Teenage visitor experience: Classification of behavioral dynamics in museums. In Proceeding of the CHI Conference on Human Factors in Computing Systems (CHI 2020), Honolulu, HI, USA, 25–30 April 2020; pp. 1–13.

35. Candello, H.; Pinhanez, C.; Pichiliani, M.; Cavalin, P.; Figueiredo, F.; Vasconcelos, M.; Carmo, H.D. The effect of audiences on the user experience with conversational interfaces in physical spaces. In Proceeding of the CHI Conference on Human Factors in Computing Systems (CHI 2019), Glasgow, Scotland, 4–9 May 2019; p. 90.

36. Machidon, O.M.; Tavčar, A.; Gams, M.; Duguleană, M. CulturalERICA: A conversational agent improving the exploration of European cultural heritage. *J. Cult. Herit.* **2020**, *41*, 152–165. [CrossRef]

37. Varitimiadis, S.; Kotis, K.; Spiliotopoulos, D.; Vassilakis, C.; Margaris, D. 'Talking' triples to museum chatbots. In *Culture and Computing*; Rauterberg, M., Ed.; Springer International Publishing: Cham, Switzerland, 2020; pp. 281–299.

38. Gaia, G.; Boiano, S.; Borda, A. Engaging museum visitors with AI: The case of chatbots. In *Museums and Digital Culture: New Perspectives and Research*; Giannini, T., Bowen, J.P., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 309–329.

39. Massaro, A.; Maritati, V.; Galiano, A. Automated Self-learning Chatbot Initially Build as a FAQs Database Information Retrieval System: Multi-level and Intelligent Universal Virtual Front-office Implementing Neural Network. *Informatica* **2018**, *42*, 515–525. [CrossRef]

40. Efraim, O.; Maraev, V.; Rodrigues, J. Boosting a rule-based chatbot using statistics and user satisfaction ratings. In *Conference on Artificial Intelligence and Natural Language*; Springer: Cham, Switzerland, 2017; pp. 27–41.

41. Hien, H.T.; Cuong, P.N.; Nam, L.N.H.; Nhung, H.L.T.K.; Thang, L.D. Intelligent assistants in higher-education environments: The FIT-EBot, a chatbot for administrative and learning support. In Proceedings of the Ninth International Symposium on Information and Communication Technology, Danang City, Viet Nam, 6–7 December 2018; pp. 69–76.

42. Sandu, N.; Gide, E. Adoption of AI-Chatbots to enhance student learning experience in higher education in India. In Proceedings of the 2019 18th International Conference on Information Technology Based Higher Education and Training (ITHET), Magdeburg, Germany, 26–27 September 2019; pp. 1–5.

43. Anne Frank House. Available online: https://www.annefrank.org/en/about-us/news-and-press/news/2017/3/21/anne-frank-house-launches-bot-messenger/ (accessed on 7 August 2021).

44. Field Museum. Available online: https://www.fieldmuseum.org/exhibitions/maximo-titanosaur?chat=open (accessed on 7 August 2021).

45. Andy CarnegieBot. Available online: https://carnegiebot.org/ (accessed on 7 August 2021).

46. Schaffer, S.; Gustke, O.; Oldemeier, J.; Reithinger, N. Towards chatbots in the museum. In Proceedings of the 2nd Workshop Mobile Access Cultural Heritage, Barcelona, Spain, 3–6 September 2018.

47. De Araújo, L.M. Hacking Cultural Heritage. Ph.D. Thesis, University of Bermen, Bremen, Germany, 2018.

48. Falco, F.D.; Vassos, S. Museum experience design: A modern storytelling methodology. *Des. J.* **2017**, *20*, S3975–S3983. [CrossRef]

49. Kang, S. Educational encounter of history and museum. *Stud. Hist. Educ.* **2012**, *16*, 7–13. [CrossRef]

50. Kolb, D.A. *The Kolb Learning Style Inventory*; Hay Resources Direct: Boston, MA, USA, 2007.

51. Schrepp, M. *User Experience Questionnaire Handbook. All You Need to Know to Apply the UEQ Successfully in your Project*; Häfkerstraße: Weyhe, Germany, 2015.

52. Othman, M.K.; Petrie, H.; Power, C. Engaging visitors in museums with technology: Scales for the measurement of visitor and multimedia guide experience. In *Human-Computer Interaction—INTERACT 2011*; Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 92–99.