

## Chatbots in Museums: Is Visitor Experience Measured?

**Kamila Štekerová**

e-mail: kamila.stekerova@uhk.cz

*Faculty of Informatics and Management, University of Hradec Králové, Hradec Králové, Czech Republic*

Štekerová, K. (2022). Chatbots in Museums: Is Visitor Experience Measured? *Czech Journal of Tourism*, 11(1-2), 14-31. DOI: 10.2478/cjot-2022-0002.

---

### Abstract

Chatbots are rapidly growing application area of conversational artificial intelligence. The aim of the paper is to explore the evaluation of user experience with chatbot applications in museums and galleries. Introduction to principles of chatbots, their creation and testing is provided. Methods of user experience evaluation are explained and the indicators that can be used to assess user experience with chatbots are listed. History and classification of museum chatbots is briefly summarized. A systematic review according to the PRISMA methodology was conducted to map the latest trends in museum chatbots' development and namely to answer two research questions: (1) What chatbots have been developed for the needs of museums and galleries? and (2) Was the visitor experience with these chatbots evaluated? The research gap in measuring visitor experience with chatbots was identified.

### Keywords

Application, chatbot, conversation, museum, tourism

**JEL classification: Z10, Z88, L83**

**Accepted: 15 October 2022**

## Introduction

The purpose of research in the field of artificial intelligence (AI) since its inception has focused, among other things, on finding ways to mimic human communication in natural language as faithfully as possible. In last years, the segment of conversational robots (chatbots) is growing in terms of all possible traceable criteria (absolute numbers of applications and users, profits, time and personnel savings for companies, share among communication channels, etc.) and across application areas (Todorov, 2022; Moran, 2022). Museums generally have space to display only about 5% of their collections at any one time, but collections are now digitised, so there is nothing to prevent their online accessibility through various applications, including chatbots. In comparison with traditional chatbot application domains (such as online marketing, telecommunications, finance, healthcare), the development of chatbots for specific use in museums, galleries and cultural institutions is creative process as it is possible to enhance chatbot functionalities with gamification, augmented reality, or virtual reality. Chatbot makes museums more attractive to the younger generation of visitors, who use digital devices and social networks almost continuously during the day.

## Theoretical basis

### Principles of Chatbots

A chatbot is a software that interacts with a user in natural language (Dahyia, 2017), usually in written form, e.g. through messaging applications, websites or mobile applications. A voicebot communicates in synthetic spoken language instead of written form. The main factors why people choose to interact with a chatbot are productivity (chatbot is available 24/7 in multiple languages), fun and curiosity (talking to a chatbot is a specific type of dialogue), and social and relational factors (the chatbot feeds the conversation, relieving the user of feelings of loneliness, the user can talk without fear of being evaluated). Chatbots are expected to play important role in customer service, being authentic voice of organization and integral part of customer relation management (CRM) strategy. The role of chatbots within customer experience and customer loyalty was studied by (Jenneboer et al., 2022). The task of the chatbot is to recognize the user's requirements or understand the question and to provide a relevant answer. From a technical point of view, the initial problem to be solved is automatic recognition of spoken speech or text, followed by the conversion of speech input to text (or vice versa) and the compilation of answers to questions (by matching keywords from the query with the content of databases of standard answers).

Simple rule-based chatbots operate with predefined scenarios. Users communicate via buttons and choosers, they answer questions, but they cannot ask, because the tree structure defining possible conversation flows is not prepared for this. If the user strays in the conversation in such a way that the chatbot does not find the keyword or scenario, it responds e.g. with standard messages ("sorry, I did

not understand", "say it in other words") and the user has the possibility to rephrase his/her question or is redirected to a human operator.

Three chatbot design archetypes (ad-hoc supporters, temporary assistants, and persistent companions) were proposed by (Nißen et al. 2022) These archetypes are specified by 22 features. Chatbots with artificial intelligence (also smartbots) apply natural language models and machine learning methods to understand sentence structures and may be equipped with the ability to learn and improve based on conversations already had with humans. The participation of human operators during its development and optimization of chatbot is essential. The more advanced the chatbot's artificial intelligence, the less additional intervention is needed. The holy grail is to make the chatbot so realistic that when the user talks to it, he is not sure whether he is talking to a machine or a real person, correspondingly to well-known Turing test.

To achieve maximum effect, it is necessary to understand and adapt to the language of target user group not only in the grammatical sense. People's ways of expressing themselves are influenced by a huge number of variable factors (such as region, age, current mood, season, weather, political beliefs, or recent experience), all of which the chatbot should respond to. The trust or distrust of the user usually depends on how the beginning of the conversation works, typically in the first three questions. According to this, the user decides whether to continue the conversation at all and finds out what to expect from the chatbot. The main step in chatbot development (after specification of project objectives and selection of platforms and communication channels) is a conversational design, which builds on three pillars:

- *The principle of cooperation* says that effective communication between two people assumes the cooperating participants (the pragmatic theory of P. Grice). This means that partners answer each other's questions naturally, relevantly, truthfully, and completely, but at the same time without an excess of unsolicited details. These are rules of human conversation that a chatbot should apply.
- *Turn-taking* means that the chatbot must pass the word to the user (or not overwhelm the user with questions that a human would not be able to respond to with sufficient alacrity). Moreover, passing the word allows to resolve ambiguities. Transitional questions and phrasing when passing the word must feel natural.
- *Principle of context* says that it is useful to consider the circumstances in which the user is in the conversation, as long as the chatbot can recognize or infer them (for example, that the user is driving and therefore does not have time for a longer discussion).

It is important to focus the social aspects of human-chatbot interactions in terms of interpersonal communication competences. From the user's perspective, numerous chatbot communication competencies are defined such as:

- *self-disclosure* (the chatbot should share personal thoughts or experience),
- *empathy* (the chatbot should show that it understands and/or sympathises with the user),
- *social relaxation* (the chatbot should appear calm, professional, and should not show concern or anxiety),
- *interaction management* (the chatbot should be able to take the floor and develop the topic of conversation on its own),
- *assertiveness* (the chatbot should defend its truth, but at the same time not interrupt or harass the user),
- *altercentrism* (the chatbot should show interest in the user, ask appropriate questions, and show emotion in a polite manner),
- *expressiveness* (the chatbot should be able to express its own feeling verbally or non-verbally using laughter, emoticons),
- *immediacy* (the chatbot should be open in communication, show active listening),
- *support* (the chatbot should not judge or evaluate the user, it should act as an equal partner),
- *environmental control* (the chatbot should meet goals, explain its limits, set adequate expectations).

Individual chatbot competencies can be easily assessed by users using a Likert scale and guiding questions (e.g., "*The chatbot seemed to know how I feel.*", "*The chatbot was able to transition smoothly between topics.*", "*I would describe the chatbot as a warm communication partner.*"). Chatbots work on all major communication platforms (Facebook Messenger, WhatsApp, Slack, Telegram, Twitter, Viber, Skype, Instagram). Until recently, development of chatbots was the domain of software developers. Due to the so-called no-code trend, visualization tools and platforms are now available (e.g. Chatfuel, Botsify, Flow XO, Motion.ai). Thanks to these platforms, chatbots are now accessible to small companies and non-profit organizations that previously could not afford them (Conejos, 2020).

An example of an open-source framework is Botpress, written in JavaScript. Its user interface can be customized by developers, with built-in AQ modules, a conversation flow editor and a natural language understanding module. Chatbots developed in Botpress can be used on Facebook Messenger, Slack, Telegram, or websites. The drawback is that Botpress does not support advanced knowledge representation (knowledge graphs), which limits the structuring of content. Knowledge graphs, on the other hand, are supported by the Rasa Framework, written in Python. Rasa consists of two components (Rasa NLU, Rasa Core) and allows to create more complex dialogues. Google's Dialogflow conversation platform provides voice and text conversational interfaces and natural language understanding. It supports most communication channels (Facebook Messenger, Slack, Skype, Kik, Telegram, Twitter, Viber, etc). It is possible to create a distributed collaborative architecture, i.e. combine multiple chatbots together. IBM's Watson Assistant cloud service uses machine learning and natural language understanding in 13 languages. IBM offers several versions of the service, including a limited free version, which is why Watson Assistant is used in research projects

and chatbot prototypes. Facebook has developed a conversational platform, Wit.ai, that supports 137 world languages, in which it can detect speaker intent, extract keywords, and recognize context.

## Testing of Chatbots

Testing chatbots is fundamentally different from testing other software applications. The difference is since a chatbot works with natural language input and must also respond to unexpected input. Chatbot testing is twofold:

- *Before launching*, the chatbot is tested in general (using basic questions and answers) and in specific terms (in terms of the application domain). The chatbot is tested for its limits, or its responses to an irrelevant or un-processable question.
- *After running the chatbot*, AB testing is performed to determine performance (by comparing two versions based on the collected data). Next, the conversational factors (whether the chatbot can initiate a conversation) and the visual aspect (the overall user interface experience) are verified.

The testing of the chatbot also includes the testing of so-called small conversations (greetings, greeting phrases influencing the user's impression). Navigation testing is important, where cases where the user skips some steps of the conversation are investigated. Emotion testing focuses on the chatbot's ability to understand the tone of the user's input (e.g. sadness, concern, joy). User acceptance testing involves conversations with different users who differ in personality traits. Automated systems are available for chatbot testing, e.g., Botium Box, Selenium, Gupshup Proxy Bot, TestFairy.

However, chatbot platforms alone do not guarantee success, which is mainly dependent on the availability of historical data usable for chatbot training. For example, to achieve a success rate of 60% in machine translation, a dictionary containing at least 30,000 phrases is required, and for deep learning in a conversation between two participants, at least 100,000 chat lines are needed, plus additional language datasets, dictionaries, etc.

The question is how chatbots contribute to the user experience and how we can evaluate this influence. The following metrics and performance indicators are commonly used in determining chatbot effectiveness (Skjuive & Brandzaeg, 2019; Goover, 2020; Roth, 2021):

- *Total number of users* (new, returning) shows the return on investment of the chatbot. Low number indicates poor placement of the chatbot on the page or technical problems (e.g. slow loading). Engaged users are people who use the chatbot repeatedly (daily, weekly). A higher number is expected e.g. for an internal help-desk chatbot that handles simple routine questions. The number of new users indicates the popularity of the chatbot. The evolution over time indicates changes in the popularity of the chatbot and derives the point at which it is appropriate to invest in more advanced chatbot features. A high number of new users with a low number of returning users may indicate that the chatbot is not beneficial to regular users. Conversely, a high number of

returning users with a lower number of new users suggests that there is a segment of the audience that has taken a liking to the chatbot.

- *Chat volume* refers to the successful conversations that the chatbot has made over a certain period of time. Low numbers are a reason to redesign the chatbot. Duration of a conversation depends on the chatbot's function. If the chatbot is to direct the customer somewhere, the conversation should be short, whereas longer conversations are expected for service support queries.
- *Total number of conversations* the chatbot handles in a day, this number grows with the user base. Average daily sessions number tells how often users initiate a call with the chatbot.
- *Completion rate* determines the success rate of the chatbot, or the achievement of its goals. Fallback rate (also Missed/Failed Utterances) tells how often the chatbot misunderstood a question due to confusing input or spelling errors and could not complete the task. Human takeover rate (also Human vs. chatbot Interaction) shows how effectively the chatbot can recognize the need to switch to a human operator.
- Bounce rate tells you how many web users have not used or seen the chatbot. Activation rate measures the chatbot's ability to initiate another call. Chatbot response time shows the speed of responses.
- If the chatbot is dealing with Frequently Asked Questions, it is useful to see from its statistics which questions it receives most often.

### Chatbots in Museums

In tourism, chatbots function in a similar way to other sectors, i.e. they provide customer service, act as personalised guides and advisors on websites, remove the problem of language barriers, provide feedback, etc. (Calvaresi & Ibrahim, 2021).

Chatbots are also already being used in the art market. For example, the ArtFinder portal introduced an AI-based virtual curator named Emma in 2016. Technically, this is the use of visual search technology, which works like the "more like this" feature on the web. Artfinder thus became the first art portal to implement automatic detection of similarities between artworks, and at the same time introduced a new mechanism for buying art. Visual search is a great help when a huge volume of artworks needs to be processed and the customer cannot articulate or keyword what they are looking for - which happens easily in the fine art industry.

Chatbots are, of course, commonly used in museum customer service, typically on museum websites, where they function similarly to, for example, e-commerce sites. This means that the chatbot helps with e.g. ticket bookings, souvenir purchases or assists in searching the digitised collections. But there is also room for more creative uses of chatbots: it is possible to connect a chatbot with gamification elements, virtual reality or augmented reality. The evolution of museum and gallery chatbots is summarised in the articles. There are many reasons to focus on this, just think of the period of museum closures during the covid pandemic, when virtual tours and online transmissions seemed

to be the solution. A categorization of the museum chatbots into five types was proposed, namely the conversational skills field and according to user engagement techniques:

- *Simple QA information chatbots* (infobots) provide information about the museum and collections with limited conversational capabilities.
- *Chatbots with conversation scenarios* where the user chooses the next course of action e.g. using menu buttons.
- *Gamification and reward chatbots* try to engage the user by playing games or offering various benefits, again with limited conversational resources.
- *Provocative conversational chatbots* control the conversation to some extent and provoke the user to ask questions.
- *Conversational chatbots* are close to human conversation, have the ability to understand the user's intent and in most cases respond correctly.

The first museum application was created in 2004 for the Heinz Nixdorf Museums Forum (Paderborn, Germany). It was a so-called avatorbot (a conversational agent called MAX) that acted as a virtual guide for visitors (Kölbl, 2018). Since 2013, the Cooper-Hewitt Smithsonian Design Museum (New York, USA) has been running a service called Object Phone, which answers questions about exhibits. As of 2016, the service is paid. Also, SFMOMA (San Francisco, USA) provides the Send Me SFMOMA SMS service through which it shares its collections with visitors. The Musée national d'art moderne, also known as the Centre Pompidou (Paris, France), launched its first Ask Mona chatbot in French and English in early 2022. The chatbot is used to browse the collections and is part of a digital strategy to make the collections accessible to as many people as possible. The chatbot is described as fun and user-friendly. The Anne Frank Museum in Amsterdam offers visitors the opportunity to see Anne Frank's house during the years she and her family hid from Nazi persecution. In 2017, the museum launched a chatbot to provide visitors with information about the exhibition. The chatbot deployed by the museum uses deep learning technology. The chatbot Máximo the Titanosaur at the Field Museum in Chicago features a dinosaur. It communicates through several channels (Facebook, Viber, WeChat, Twitter, etc.). He comes across as friendly, able to answer all sorts of questions in a human-like manner. In addition to the use of language understanding and machine learning techniques, extensive user testing and dialogue redesign was conducted. The museum in Rio de Janeiro, Brazil, runs an IRIS chatbot that was developed using the IBM Watson framework and communicates in English and Portuguese and in spoken language; there is a version with a text input interface for visitors with hearing impairments.

## Methodology

A review of museum chatbots according to the principles of the PRISMA (Page et al., 2021) methodology was conducted to map the latest trends in the development of museum chatbots and to answer two research questions:

Q1: A designer's perspective: What chatbots have been developed for the needs of museums and galleries?

Q2: User perspective: Was the visitor experience with these chatbots evaluated?

Scientific databases Web of Science, Scopus and Science Direct were used for the search. English keywords searched were *chatbot* and *museum*. Preliminary research has shown that no refinement of the query with additional keywords (*conversational agent, virtual assistant, application, case study, example, gallery* etc.) generally returns additional relevant results. The search was limited to publications published in 2018 and later, with a minimum of older papers given the newness of the technology. A summary of the queries is presented in Table I.

**Table 1** Search in Databases (Source: author, 2022)

Database	Query	Number of Results
Web of Science	Keywords: chatbot AND museum (All Fields) Refined By: Document Types: Proceedings Papers or Articles. Publication Years: 2018–2022	7
Science Direct	Keywords: chatbot AND museum Years: 2018-2022 Article type: Research articles, Short communication Subject area: Computer Science, Arts and Humanities	17
Scopus	Keywords: chatbot AND museum Years: 2018-2022 Article type: Conference proceedings, Conference Review, Article	14

The second step was to evaluate the abstracts. Articles that did not meet other required criteria were excluded: either they did not correspond thematically to the application domain, or they did not describe the chatbot application, or the full text of the article was not available. The remaining 11 papers were analysed to answer the research questions.

## Results

The content of 11 papers is briefly summarised in following paragraphs.

### **Application 1: MUSU** (Jaramillo et al., 2018)

*Objective:* A prototype of a conversational application called MUSU is designed for museum visitors to enhance the interactivity of exhibitions.

*Technology:* The app can actively converse and ask questions. Communication is possible in written and spoken form. IBM Watson technology is used, a prototype has been developed for the iOS operating system, and an Android version is planned.

*Testing:* The authors conducted a survey among two groups of visitors (children 4-7 years old, adults 24-55 years old) to find out the most common questions visitors ask in connection with the exhibited works. Differences between visitor groups were identified. The chatbot was not subject to testing.

### **Application 2: Le Case Museo di Milano** (Gaia, Boiano & Borda 2019)

*Objective:* Le Case Museo di Milano includes four houses from different historical periods. These are not classic exhibitions with systematically arranged exhibits with captions, but inhabited houses, preserved as their owners arranged them. The initial question addressed by authors was how to make the exhibitions attractive to teenage audience.

*Technology:* A mobile application was designed, combining a chatbot with gamification elements. The chatbot game is available via Facebook Messenger and Telegram Messenger, the aim of the game was to search for clues in the museum's collections and solve a mystery (the game features a girl who asks visitors to help her defeat a mysterious astronomer working in the court of a 15th century Milanese duke). The chatbot has two ways of interacting with the user: (1) it asks open-ended questions (this is implementation-intensive, but it has more effect on the visitors' emotions), (2) it asks multiple-choice questions where the user selects using buttons (this keeps the game flowing).

*Testing:* 80 students aged 16-18 participated in the pilot testing. The results were as follows: 90% of the students were able to complete the game; 30% had connection problems; 34% were concerned about the security of their data; 88% rated the length of the game as appropriate; 72% rated the game as very fun; 66% found it a useful learning tool when used with another student or in a small group (i.e., users preferred using the chatbot in a group to interacting with it individually).

120 users participated in the second testing. It was found that teenagers readily accept a game that acts as a guide in a museum. However, some pointed out that if they had not visited the museum beforehand, the game made it difficult for them to grasp the overall context of the museum on their first visit. The solution is to offer the game only to returning visitors, or to offer a short tour of the museum before the game begins.

### **Application 3: ChatÇat** (Roussou & Perry, 2019)

*Objective:* Authors focus on persuasive chatbots and address the question of whether acting as a chatbot can make people change their perceptions and think critically. They therefore propose a provocative chatbot of conviction to spark conversations about complex, emotionally challenging or controversial topics (e.g. gender issues, power sharing) when talking about the UNESCO World Heritage site of Çatalhöyük, a Neolithic archaeological site in Turkey.

*Technology:* The first step was the development of a traditional infobot working with the knowledge base of the topic. A basic version was developed based on live chat sessions between experts and end users. The chatbot was then retrofitted with more complex reflective and emotional components, and the basis for its development was based on posts, comments and shares from the Çatalhöyük Research Project site's Facebook page (over 10,000 users follow the site as of April 2022) and other live chats with the public. In this way, the topics of interest to visitors (including burials, wall paintings, the archaeological process, the site's surroundings, and the significance and chronology of the plaster and figurines) were identified. A chatbot ChatÇat was created for Facebook Messenger. The conversation model uses a pattern that is similar to the so-called Socratic dialogue. The chatbot performs a simple form of interrogation of sorts, where it asks questions and tries to trace the user's reasoning while maintaining control over the direction and structure of the conversation. The chatbot begins by asking a question, making a bold statement of its own or expressing an opinion, to which the user is forced to respond. Further dialogue takes place according to the chosen answer, which has three possible forms (yes, no, ambiguous answer). The conversation proceeds according to a given design pattern:

- Provocation, i.e. the chatbot declares an opinion or asks a provocative question.
- The user responds positively (yes), negatively (no) or ambiguously (everything else).
- An exchange of ideas takes place (2-3 interactions).
- In the middle of the conversation, the chatbot tests the user's beliefs.
- The conversation ends with a final statement of belief.

*Testing:* Testing of the chatbot has so far only been conducted with a small, unrepresentative sample of users (2 men, 7 women). The functionality of the proposed conversation model was confirmed, users were intrigued by the provocative questions and radical views of the chatbot, which additionally revealed their own preconceived ideas about the past. The resulting impression was that the chatbot took the initiative in the conversation and forced the user to assert his or her opinion.

#### **Application 4: Chatbot with Mixed Reality (Ali, Le & Kim, 2019)**

*Objective:* Multimodal solution was designed for museums, botanical gardens and similar places.

*Technology:* A combination of chatbot and wearable mixed reality supports spatial mapping. A virtual character (i.e. a domain-specific 3D chatbot) acts as a guide and together with an object recognition system aims to provide a very realistic, immersive experience for the visitor. The design and implementation of a virtual agent for wearable augmented reality is described in terms of technical implementation. Gaze expression techniques (using the face of a 3D virtual character) in conjunction with high-level speech processing are used to achieve natural-looking interaction. The chatbot provides information and is also equipped for general conversation on an open topic. It also has an emotion processing component (sentiment engine) for analysing responses and assigning them to a so-called sentiment class. The authors present a sample scenario, which is a tour of a botanical garden. The user can interact with a virtual character, i.e. ask questions about the plants and discuss them. The scenario works with a virtual character with a realistic appearance, as well as a knowledge base about plants and an object recognition system.

*Testing:* The purpose of the testing was to check the technical functioning of the application, especially the speed of response and therefore the smoothness of the experience, also the reliability of the object recognition method used. The user experience is not discussed in the article.

#### **Application 5: Leonardo da Vinci's Machines** (Caggianese & De Pietro, 2020)

*Objective:* This study focuses on modern visualization technologies and their application in museums. It presents an interactive holographic stand that presents the art of Leonardo da Vinci, specifically his machines, to users. The machines can be interactively manipulated through a touchless user interface.

*Technology:* The system uses a display technique that creates a holographic effect: 3D content is projected onto a semi-transparent and semi-reflective surface, positioned at a 45° angle from the projection source (55-inch screen). The design includes a webcam and a directional microphone, which are used to recognize the active visitor and obtain input. A Leap Motion controller is used for touchless interaction with the projection panel. The application includes a chatbot capable of conducting dialogue in natural language. The Google Speech interface is used to evaluate the spoken input and convert it to text, and IBM's Watson cloud-based natural language classifier is used for intent recognition. The classifier was trained using a prepared dataset (list of questions with associated intent).

*Testing:* An analysis of user interactions with the system was performed to see how the total time spent with the application, the number and types of queries, or the time spent manipulating objects varied across user groups. The testing involved 164 users aged 6-72 years who interacted with the system for at least 30 seconds on at least one day. Users were not given detailed instructions, only informed that the system was capable of interactively visualizing holograms of five Leonardo machines and answering their questions in natural language. The resulting findings are presented in the paper with detailed statistics.

#### **Application 6: AIMuBot** (Zhou, Sinha & Liu, 2020)

*Objective:* The article An AI chatbot for the museum based on user Interaction over a knowledge base describes the technical design of the chatbot called AIMuBot.

*Technology:* The chatbot connects a knowledge base with an interactive information retrieval system. Communication is possible in spoken language, and a graphical interface is available to specify questions and visualize answers. The system consists of four modules:

- A graphical user interface helps users formulate questions and displays answers.
- The extractor module identifies the query entities from the natural language input.
- The query refinement module asks the user additional questions.
- The visualizer module processes the refined query, i.e., interacts with the knowledge base and returns the result.

The AIMuBot demo works with the freely available dataset of the Metropolitan Museum of Art in New York, from which 14406 exhibits related to Chinese culture were used for the demo.

*Testing:* User testing is not mentioned in the paper.

#### **Application 7: CHAT-Bot** (Casillo et al., 2020)

*Objective:* The authors of the article CHAT-Bot: A cultural heritage aware teller-bot for supporting touristic experiences is to present a chatbot based on the principle of a context-aware system that can recommend content and services to visitors of the Herculaneum and Pompeii sites in Southern Italy.

*Technology:* The architecture of the CHAT-Bot is based on several main modules. The storytelling module is designed to convey an experience to the user through a personalized story. In order to respect the differences between visitors and to achieve a sense of uniqueness of the experience, a number of

scenarios and itineraries have been created. The context-based manager deals with the representation of all possible contexts of use through a context tree. The chatbot can be interfaced with external resources (e.g. ticket booking systems). Other modules implement the actual dialogue with the user, track the timing of interactions, etc. The core of the application is an inference engine that performs text analysis and context extraction.

*Testing:* Testing of the chatbot involved 3150 users aged 18-62 who rated five aspects of the chatbot using a Likert scale:

- A (Recommendations) - whether the chatbot recommended services and content in line with the user's needs, preferences and current context.
- B (Conversation) - whether the dialogue with the chatbot went smoothly and whether the chatbot understood the user's intent.
- C (Presentation) - the information provided here was comprehensive.
- D (Usability) - whether the chatbot interface was user-friendly and the response was fast enough.
- E (Future developments) - whether it would be appropriate to invite other people (e.g. family members travelling together) to participate in the communication.

The possible answers were "I totally disagree" - TD, "I disagree" - D, "Undecided" - U, "I agree" - A, "I totally agree" - TA. The results of the evaluation are presented in the form of tables.

#### **Application 8: National Museum of Korea** (Noh & Hong, 2021)

*Objective:* The authors of Designing Reenacted Chatbots to Enhance Museum Experience investigated the usability of three chatbot models in history education with respect to users' individual learning styles. This is an e-learning rather than a direct museum application. Therefore, in the introduction of the article, the authors also list the educational applications of chatbots. The chatbots were developed in collaboration with the National Museum of Korea. The topic of the conversation was the Silla Kingdom (from the 1st millennium AD), an important period for Korean history.

*Technology:* To create the knowledge base, frequently asked questions and answers were extracted from web search engines regarding the landmarks, cultural and political background, and historical figures from the Silla period. Based on the results, knowledge structuring was conducted on sub-topics, namely (1) personalities of the Silla Kingdom; (2) lifestyle; (3) monuments; (4) specific knowledge about King Jinheung: personal information, historical events, political system introduced by him. All contents, question and answer pairs, and the design of chatbot models were revised by the museum curator. Three versions of the chatbot were created:

- A rule-based chatbot, a Q/A system or simple Q&A application) provides objective answers from a third-party position.
- A chatbot embodying a guide or teacher, which is the model of information delivery that visitors typically expect in a museum or gallery. The chatbot expresses itself in modern language and speaks from an observer's perspective.
- A chatbot embodying a historical figure (a king) comes from the past, has an appropriate costume and demeanour, speaks in archaic language and presents information from the participant's perspective.

Chatbot responses were both textual and spoken. Audio responses were generated using a podcast program. For the so-called embodied chatbots (teacher, king), a 3D visualization with five movements (e.g., bowing, pointing, shaking hands) was designed; the historical character also had five facial expressions (e.g., joy, surprise, anger).

*Testing:* The authors asked: (1) What are the factors of chatbot design that influence users' interaction and experience in the museum? and (2) What are the specific relationships between visitors' learning styles and their museum experience? Thirty-four respondents in their 20s participated in the testing. Respondents had sequential conversations with individual chatbots and completed questionnaires; a gaze tracker was also used to provide an in-depth analysis of respondents' behaviour during the conversation. A User Experience Questionnaire was used to determine the overall impression and general evaluation of the chatbots. Six sub-factors were evaluated:

- chatbot attractiveness,
- clarity of communication with the chatbot,
- effectiveness of solving a task or answering a question,
- reliability, or the user's feeling that they can control the interaction with the chatbot,
- incentive, or inducing motivation to continue using the chatbot,
- the overall novelty of the application.

The second evaluation focused on the impact of chatbots on the museum experience, using the Museum Experience Scale (MES) survey. The evaluation consists of four components (engagement, knowledge/learning, meaningful experience, emotional attachment), each component is assessed using five questions. The results of the testing are summarized in graphs and text in the paper. The authors conclude that the concepts of embodiment and re-enactment in chatbot models improve the user experience and contribute positively to learning. Furthermore, the authors conclude that there is no perfect chatbot model applicable to every situation. Objective, factual, expert-verified information is easier to provide through a rule-based (Q/A) chatbot, which is not flexible in generating responses.

#### **Application 9: CHIM** (Schaffer et al. 2021 ; Shaffer et al. 2018)

*Objective:* A prototype of chatbot was created for the Städel Museum in Frankfurt am Main.

*Technology:* Visitors can ask the chatbot questions about the exhibits, the answers are available in different formats (text, sound, image, video). The paper describes the technical solution (categorization of visitors' questions, natural language processing mechanisms used) and explains future steps (implementation of dialogue management, improvement of natural language processing strategies by incorporating additional answers for selected exhibits, final versions of the interface).

Visitor queries received through the website were manually annotated and categorized into 12 categories. 2357 queries were processed. The categories were designed to cover particular aspects of the exhibits (facts about the exhibits, details about the life and work of the artists of the works, questions about the technical execution of the artworks, questions about artistic styles, contextual questions about historical, political or social context, requests for an interpretation of the deeper meaning of the work or the artist's intention, questions that go beyond the usual categories, etc.). When it was found that more than half of the questions related to the meaning of the artworks, the categorisation was modified and four new categories were introduced (content - questions about what is depicted in the artwork, including details; subject matter - questions about the original subjects,

people or events; reaction - questions about the reaction the work evokes in the audience; provenance, i.e. the chronology of ownership or location of the works).

*Testing:* Testing or evaluation of the chatbot application is not the focus of this paper.

#### **Application IO: MuBot** (Varitimiadis et al. 2020)

*Objective:* The article "Talking" Triples to Museum Chatbots presents the use of intelligent chatbots for museums in conjunction with Knowledge Graphs (KG) technology to represent and structure knowledge about exhibits. It describes – In the authors' opinion – the first experimental intelligent AI chatbot based on KG and ontology. The chatbot is called MuBot and was developed for the needs of the Archaeological Museum in Heraklema, Greece.

*Technology:* The architecture of MuBot consists of three main components:

- Knowledge base, which uses Semantic Web technology (RDF, SPARQL query language and OWL ontology) for knowledge representation, querying and reasoning.
- A natural language component that translates natural language user input into RDF/SPARQL,
- A natural language generation component that generates natural language responses that affect the user.

MuBot should be able to conduct conversations from the position of a museum guide as well as from the position of a talking exhibit. The resulting vision of a smart museum should therefore offer both classic exhibits and their virtual counterparts. The authors describe the process of creating a chatbot as follows:

- First, the purpose of the chatbot was defined, a detailed documentation (knowledge base) of the selected exhibit (the figure of the so-called Snake Goddess, which is one of the popular exhibits) was created, and a search of external resources and other available content was conducted.
- A chatbot for the experimental exhibit was created using the Dialogflow platform. A knowledge graph captured information about the chronology, material, and purpose of the figurine, and questions and answers related to each content category were added.
- The next step was to create the chatbot infrastructure and refine it with machine learning methods.
- An API was created and the MuBot was linked to a user-friendly website.

*Testing:* Testing was carried out with the participation of experts (3 mobile application programmers, 2 museum subject matter experts). Detailed user experience tests are not presented in the paper.

#### **Application II: Collaborative Chatbots** (Varitimiadis et al. 2021)

*Objective:* The team that created the MuBot (Varitimiadis et al. 2020) application has also published more recent work in which the importance of knowledge graphs is emphasized and a distributed architecture of multiple chatbots (collaborative chatbots) is outlined, an important trend towards which research in machine learning and language processing is moving. A knowledge graph describes real-world entities and their relationships in the form of an oriented graph. The integration of knowledge graphs is typical of hybrid conversational artificial intelligence. It allows the creation of more elaborate dialogues, it is possible to integrate different knowledge bases, e.g., extracted from the Semantic Web, so that the conversation does not have to stick only to prepared topics in terms of content. It is thanks to knowledge graphs that different chatbots can communicate with each other.

*Technology:* This paper presents a proposal for a distributed collaborative architecture composed of multiple museum chatbots. Each chatbot has its own topic (domain) of conversation, its own knowledge base and can collaborate with other chatbots to form a network. Thus, chatbots provide

autonomous conversational services to each other based on their skills and can also collaborate to provide information to the museum visitor. The Chatbot Service can be accessed through different channels (e.g. Facebook, WhatsApp) and devices (computer, tablet, phone). The knowledge index storage mechanism recognizes the user's intent and reaches out to a suitable chatbot that can provide the required information. A natural language processing component will convert the user input into the necessary formalism. The knowledge graph component (using Semantic Web technology) handles the representation, linking, reasoning of answers. The natural language generation component returns answers in a user-friendly form.

*Testing:* No testing or user evaluation is done in this article, it is an ongoing research.

## Discussion

### Answer to QI: Designer's Perspective

Chatbot technologies are increasingly penetrating museums and galleries. The number of applications has been growing over the last four to five years, undoubtedly related to the rising popularity of smartphones and the growth of social networking, the wider range of chatbot implementation platforms, and the progressive digitisation of museum collections. Of course, the 2020-2021 period of across-the-board lockdowns has also increased interest in all sorts of new technologies. In particular, descriptions of prototypes and projects in progress were found in research databases. Developments range from different variants of virtual guides and databases of prepared questions and answers to more complex architectures (AIMuBot, MuBot) and embodied chatbots (Le Case Museo, National Museum of Korea) to chatbots working with emotions (ChatÇat) and distributed collaborative architectures. The articles found were mostly published at AI conferences, where their authors mainly presented technical solutions. Researchers focused on structuring knowledge bases, on proposing variants of conversational design and on achieving a realistic impression. Attention is paid to structuring question bases and answering advanced knowledge representation methods and user context recognition.

IBM Watson (MUSU, Leonardo da Vinci's Machines) technology is commonly used to implement conversations. The choice of platform is based on the habits and expectations of the target users. The user should not be forced to install a new mobile application just for the need of visiting the museum. Therefore, chatbots integrate into multiple platforms simultaneously (Facebook Messenger, Telegram Messenger, WhatsApp Snapchat predominate). The risk is the dependence on the development platform for which the application is created. For example, Facebook has seen an exodus of younger users in recent years, while other social networks are growing; overall, development is rapid and hard to predict, making it difficult to make decisions when choosing a platform for chatbot development. The solution is multi-platform development and continuous market research.

The purely technical issues relate to the need for a reliable Wi-Fi connection to maintain constant contact with the chatbot application, which can be more difficult in historic museum buildings and grounds. The speed and response quality of the chatbot also needs to be addressed. For

a chatbot to succeed, it must not impose technical barriers on the user (e.g., forcing the user to make complex settings or install additional applications).

The combination of chatbots with gamification features is expected to appeal mainly to younger (child) audiences. One possibility is to develop and create characters with feelings and personalities, so the chatbot can adapt to the needs and age of the audience. The gamification approach (Le Case Museo), which relies on storytelling, appeals to emotions and assumes the user is interested in competing or cooperating. If the chatbot is linked to a game, its scenario and timeframe (pace of the game, its duration) must be addressed so that the game corresponds to the time spent in the museum. The game can be designed for regular visitors or, on the contrary, for random visitors. There is also the question of the preparation phase (e.g. what the visitor should know before visiting the museum, or before playing the game).

The sophistication of the chatbot depends on the budget and other possibilities of the museum, as technologies are constantly changing. While the chatbot itself is a low-cost solution, connecting it to other technologies (e.g. virtual reality, augmented reality, holograms) and/or creating a rich knowledge base with multimedia is much more expensive both in the creation phase and later in operation and maintenance.

#### Answer to Q2: User's Perspective

User involvement in development and testing is characteristic of conversational artificial intelligence and, more importantly, necessary. In the case of museum applications, various sources of visitor data, mock conversations between visitors and experts, or social networks are exploited to best structure the content that the chatbot will present and to ensure that the chatbot meets the preferences of the target group in all aspects.

Testing is done either with large groups of respondents who give their opinions using a limited scoring scale, or through in-depth interviews, measurements and surveys with a few users to get richer feedback. Most of the articles included in the search tested technical aspects of chatbots or specific user behaviour (National Museum of Korea – learning styles). User experience and user experience were rather marginally investigated, using small samples (ChatCat users). The purpose of exploring visitor queries was to better structure the knowledge base (CHIM, MuBot).

Evaluation of user (customer) experience using standard metrics was not conducted in any of the research. Such an evaluation would only make sense after a longer period of chatbot operation. This would give the institution an idea of its benefits and risks and could guide its further development and use. Chatbots are generally well received by participants in the testing. Children and adolescents are the most difficult target group to reach, because on the one hand they are well versed in the technology, on the other hand they are critical of both museums and apps and tend to test their limits.

## Conclusion

Conversational artificial intelligence represented by chatbots is developing rapidly today. In business, it brings a new, mostly positive experience to customers, and for companies it is a way to streamline their own operations. Chatbots are better accepted by younger users (millennials, generation Z), who hardly put down their mobile phones during the day and at certain moments it is not even crucial for them whether they are communicating with a live human or a chatbot. The older generation is more cautious, but on the other hand they represent a huge target group. Here it must be admitted that chatbots have two faces: on the one hand, they appear to be attractive, user-friendly solutions that bring only benefits to companies and customers, on the other hand, some users have mixed feelings about chatbots, or concerns about their privacy and data.

There is an interest in museums, galleries and similar institutions to use chatbots to increase the interactivity and attractiveness of exhibitions. Museums have the advantage of being established and trusted institutions where research projects can be carried out, including research on conversational AI. As museums' budgets tend to be limited compared to those of the commercial sector, the availability of chatbot platforms that make application development more affordable is crucial for chatbot development. Systematic research has shown that in general, chatbot applications in museums are still at the prototype stage, typically integrating multiple new technologies and using them creatively (linking chatbots to virtual, augmented or mixed reality, incorporating gamification elements, holograms). The customer (or visitor, user) experience with chatbots is not accurately assessed by the usual metrics for measuring chatbots, due to the fact that applications have not been deployed in use for long enough and therefore the need for such measurement has not yet arisen.

The trend, as in other conversational AI applications, is to improve the level of conversation of the chatbot, i.e. to move it from a virtual guide to a full-fledged partner that will dialogue with the user and understand the context. The evolution is towards more open chatbots that act on the emotions of users. The effect of the experience depends on the methods used to recognise and capture the emotional states of the user, the design of the responses to these states and the overall design of the user interface. The integration of the chatbot into the overall concept of the exhibition or museum is essential. The use of chatbots must be in line with the institution's strategies towards the visiting public. There is already a shift from traditional exhibitions and displays to so-called smart museums, where digital counterparts with links to social networks operate alongside traditional exhibits. The visitor has the opportunity to interact with the exhibitions outside opening hours and in a more varied way than in the case of traditional exhibitions. Understandably, the chatbot must not distract the visitor from the museum itself towards applications and mobile devices. Challenges for further development include ensuring easy access to the apps for e.g. disabled visitors, app security, data protection and user privacy.

## Acknowledgment

The financial support of the Specific Research Project “Information and knowledge management and cognitive science in tourism” of YW University is gratefully acknowledged.

## References

- Ali, G., Le, H. Q., Kim, J. et al. (2019). Design of Seamless Multi-modal Interaction Framework for Intelligent Virtual Agents in Wearable Mixed Reality Environment. *In Proceedings of the 32nd International Conference on Computer Animation and Social Agents (CASA '19)*, pp 47–52. New York, USA: Association for Computing Machinery.
- Borda, A., Gaia, G., & Bioano, S. (2018). *Chatbots in museums: hype or opportunity?* [2022-04-10] Retrieved from [https://www.researchgate.net/publication/324439065\\_Chatbots\\_in\\_museums\\_hype\\_or\\_opportunity](https://www.researchgate.net/publication/324439065_Chatbots_in_museums_hype_or_opportunity).
- Caggianese, G., De Pietro, G. et al. (2020). Discovering Leonardo with Artificial Intelligence and Holograms: A User Study. *Pattern Recognition Letters*, 131. DOI: 10.1016/j.patrec.2020.01.006.
- Cassilo, M. et al. (2020). CHAT-Bot: A cultural heritage aware teller-bot for supporting touristic experiences. *Pattern Recognition Letters*, 131, 234-243. DOI: 10.1016/j.patrec.2020.01.003.
- Calvaresi, D., Ibrahim, A. et al. (2021). The Evolution of Chatbots in Tourism: A Systematic Literature Review. In Wörndl, W., Koo, C., Stienmetz, J.L. (Eds), *Information and Communication Technologies in Tourism 2021*, pp 3-16. Springer, Cham. DOI: 10.1007/978-3-030-65785-7\_1.
- Conejos, F. (2020). *How Can Chatbots Improve Customer Experience?* [2022-02-08] Retrieved from <https://www.processmaker.com/blog/how-chatbots-improve-customer-experience/>.
- Dahiya, M. (2017). A tool of conversation: Chatbot. *International Journal of Computer Sciences and Engineering*, 5(5), 158–161.
- Gaia, G., Boiano, S., Borda, A. (2019) Engaging Museum Visitors with AI: The Case of Chatbots. In Giannini, T., Bowen, J. (Eds), *Museums and Digital Culture. Springer Series on Cultural Computing*, pp 309-329. Springer, Cham.
- Goover, K. (2020). *16 KPIs to Measure Chatbot Effectiveness*. [2022-03-25] Retrieved from <https://www.lyntonweb.com/inbound-marketing-blog/16-kpis-to-measure-chatbot-effectiveness>.
- Jaramillo, J.C., Occhiuto, D., Garzotto, F. (2018). Artworks' Features Discovery Through Engaging Conversations for Children. *IOP Conference Series: Materials Science and Engineering*, 364 (1).
- Jenneboer, L.; Herrando, C.; Constantinides, E. (2022) The Impact of Chatbots on Customer Loyalty: A Systematic Literature Review. *Journal of Theoretical and Applied Electronic Commerce Research*, 17, 212-229. DOI: 10.3390/jtaer17010011.
- Kölbl, A. (2018). *Talk to me! Chatbots in Museums: A Chronological Overview*. [online, 2022-04-15]. Retrieved from <https://zkm.de/en/talk-to-me-chatbots-in-museums>.
- Moran, M. (2022). *25 Top Chatbot Statistics For 2022: Usage, Demographics Trends*. [online, 2022-04-10]. Retrieved from <https://startupbonsai.com/chatbot-statistics/>.
- Noh, Y.G., Hong, J.H. (2021) Designing Reenacted Chatbots to Enhance Museum Experience. *Applied Sciences*, 11(16), 7420.
- Nißen, M. et al. (2022). See you soon again, chatbot? A design taxonomy to characterize user-chatbot relationships with different time horizons. *Computers in Human Behavior*, 127, 107043. DOI: 10.1016/j.chb.2021.107043.

- Page, M.J. et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372 (71). DOI: 10.1136/bmj.n71.
- Roth, L. (2021). *6 essential KPIs to Track Through Your Chatbot Analytics*. [online, 2022-04-15]. Retrieved from <https://blog.alexa.com/chatbot-analytics/>.
- Roussou, M., Perry, S. et al. (2019). Transformation through Provocation? *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*, 627, 1–13.
- Schaffer, S. et al. (2022). Questions and Answers: Important Steps to Let AI Chatbots Answer Questions in the Museum. In Wölfel, M., Bernhardt, J. Thiel, S. (Eds), *ArtsIT, Interactivity and Game Creation. Creative Heritage. New Perspectives from Media Arts and Artificial Intelligence. 10th EAI International Conference, ArtsIT 2021, Virtual Event, December 2-3, 2021, Proceedings*. Springer.
- Schatffer, S. et al. (2018). Towards chatbots in the museum. *CEUR Workshop Proceedings*, 2176.
- Skjuve, M., Brandzaeg, P.B. (2019). Measuring User Experience in Chatbots: An Approach to Interpersonal Communication Competence. *Internet Science. LNCS*, 11551, Springer.
- Todorov, G. (2022). *117 Amazing Chatbot Statistics 2022*. [online, 2022-04-15]. Retrieved from <https://thrivemyway.com/chatbot-statistics/>.
- Varitimiadis, S., Kotis, K. et al. (2021) Graph-Based Conversational AI: Towards a Distributed and Collaborative Multi-Chatbot Approach for Museums. *Applied Science*, 11, 9160.
- Varitimiadis, S., Kotis, K. et al. (2020). “Talking” Triples to Museum Chatbots. 2020. Culture and Computing. HCII 2020. *Lecture Notes in Computer Science*, 12215. DOI: 10.3390/app11199160.
- Zhou, C., Sinha, B., Liu, M. (2020). An AI chatbot for the museum based on user Interaction over a knowledge base. In *AIAM2020: Proceedings of the 2nd International Conference on Artificial Intelligence and Advanced Manufacture*, pp 54–58. DOI: 10.1145/3421766.3421888.