Exploring Beyond the Exhibits
Creating Knowledge for Social Robots in Public Spaces

Ana Müller, Michael Schiffmann, Anke Neumeister, Anja Richert[1]

Technological advancements are integrating social robots into public spaces. The scientific community has consequently become aware that relying on laboratory studies may not be the best approach to developing (social) robots, particularly when they are meant to interact with humans. Studies of human-robot interaction in real-world settings, such as museums, are deemed more beneficial in understanding the complexities of these interactions (Sabanovic et al. 2006). Our goal in conducting this study was thus to better understand the requirements with respect to social robots in public spaces and to make them as representative of real-world conditions as possible by drawing on actual users, systems, and environments for real-world tasks. Our analysis of user utterances offers insights into identifying user expectations as well as system limitations and weaknesses that can be addressed through further development and training.

Related Work

Several attempts have been made to deploy robots in museums. These robots are used to greet visitors, provide information or navigation, and augment exhibitions with additional content and services (Villaespesa 2021; Willeke et al. 2001; Cantucci/Falcone 2022). For example, in 2018, the Smithsonian Museum used Pepper (Softbank Robotics) robots to translate, guide visitors, and teach coding, while at the Akron Museum a robot tour guide named Dot, which shares art information and

1 In this research, the projects SKILLED—Socioempathic AI-Based Dialogues (BMBF) and (Digital) Meer erleben (Experiencing the [Digital] Ocean, BKM) collaborated. The research project SKILLED at the TH Köln – University of Applied Science, Cologne Cobots Lab is funded by the Federal Ministry of Education and Research of Germany (BMBF) in the framework of FH-Kooperativ 2–2019 (project number 13FH504KX9). The project (Digital) MEER erleben is funded by the German Federal Government Commissioner for Culture and the Media (BKM) in the framework of museum4punkt0. We thank all the collaborators for their assistance and contributions.

asks visitors questions, was evaluated (Smithsonian, 2018; Villaespesa, 2021). Even though the use of Pepper robots at the Smithsonian Museum may give the impression that the robot performed various tasks, there were multiple robots, each for one respective application. For instance, a system designed to greet visitors may not have extensive knowledge about exhibitions, while a navigation robot may have detailed knowledge of the contents of the museum and navigating through it but less knowledge relating to service issues or small talk with visitors.

As long since pointed out by Skantze/Al Moubayed/Gustafson et al. (2012), developing social robots and situated dialogues in public settings with real users is a challenging goal for several reasons. Dialogue systems, also known as conversational systems, can be categorized into three types: 1) goal-based systems, designed to complete specific user tasks such as scheduling appointments, and typically gather information through questions until the task is fulfilled (Gao et al. 2018; McTear/Callejas/Griol al. 2016); 2) chatbots, designed for casual conversation on open-ended topics, and trained end-to-end using large datasets of dialogue examples (ibid.); and 3) question-answering systems, which focus on answering a wide range of questions from a knowledge base, with the emphasis on information retrieval (Dimitrakis et al. 2020).

In social robots such as Pepper and Nao (Softbank Robotics), or Furhat (Furhat Robotics), simple goal-based dialogue systems are commonly used due to the ease of controlling the knowledge and dialogue (Foster 2019). This nonetheless requires every spoken interaction with the robots to be handcrafted and a dialogue policy to be programmed. User utterances during a conversation are processed by a natural language processing unit, which identifies the intended meaning of the user's words. But deploying social robots effectively in real-world settings requires extensive preparation, and creating knowledge for a robot without a specific area of operation is challenging. The difficulty lies in having to handcraft both potential user utterances and robot responses. Depending on the natural language processing software used, practitioner guides recommend creating 20 to 60 training phrases per intent, but not more than 80 utterances, so as to avoid overlapping training phrases that would decrease the precision of the natural language processing (LivePerson Inc. 2022).

Despite advances in social robot technology, the existing guidelines for their development do not fully address the importance of conversational abilities or the specific topics that robots should be equipped to discuss. Further research and development are needed to ensure that these systems are well-prepared for public spaces (Mintrom et al. 2022; Tian/Carreno-Medrano/Sumartojo et al. 2020). This is nonetheless particularly challenging because of varying user expectations, and some users may attempt to test the intelligence and capabilities of the system, such as its ability to understand natural language, process text, and perform specific tasks like counting or sentiment analysis, driven by motivations other than a need for information and service, such as curiosity about the system's abilities.

Besides the challenges posed by producing knowledge and acknowledging the diversity and varying demands of different users, additional phenomena of human-robot interaction must also be addressed and recognized from the perspectives of various users and stakeholders. One of these aspects is undoubtedly the phenomenon of anthropomorphism. Based on the studies conducted on human-robot interactions, it is evident that people tend to attribute humanlike qualities to nonhuman entities (Reeves/Nass 1996; Epley/Waytz/Cacioppo et al. 2007; Nass/Brave 2005). The computers are social actors (CASA) paradigm demonstrates that advancements in technology continue to encourage the personification of objects in human-machine interactions. Anthropomorphism also has significant implications for how humans perceive and interact with (social) robots, regardless of the embodiment, interface, or use case of the systems involved, including but not limited to chatbots, digital assistants, smart homes, and social robots interacting with museum visitors (ibid.; Ivanov/Webster/Berezina et al. 2017).

Methodology

The OZEANEUM in Stralsund is a natural history museum that is part of the German Oceanographic Museum Foundation. With an exhibition area of 8,700 square meters and 50 aquariums, the museum presents a journey through the underwater world of the northern seas. Along with three other locations of the German Oceanographic Museum Foundation, the OZEANEUM is one of the most visited museums in Germany. The museum is aimed at locals and guests alike and targets all age groups and social classes. Its events, publications, and research projects also make the museum interesting for professional audiences. Families with children are one of the OZEANEUM's core target groups, and their share consistently exceeds 50 per cent. The presence of particular target groups varies considerably depending on the season. The visitor structure of the OZEANEUM is strongly influenced by tourism, especially during the high season. High visitor numbers at particular times, such as on bad weather days in summer, lead to limitations on the visitor experience and the entire visitor stay. Under such conditions, it can be difficult to find lifts, toilets, and certain exhibition and aquarium areas, and the museum's visitor services are overwhelmed. At times, it can feel similar to being at a crowded train station or airport. One solution to this issue could be to provide support by means of a social robot.

Figure 1: Interaction with the social robot Furhat at the OZEANEUM in Stralsund in March 2022. Source: Anke Neumeister, German Oceanographic Museum, CC-BY-SA.

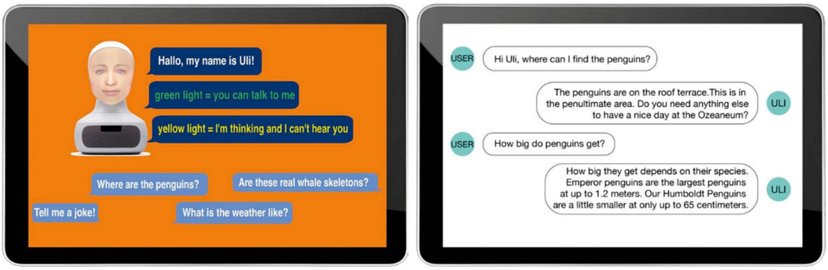
The Research System

The robotic system used in the experiments consists of a 41-centimeter-tall Furhat robot head (see fig. 1) with an illuminated mask system from Furhat Robotics and a dialogue system running on a backend server, developed at the University of Applied Sciences Cologne as part of the research project SKILLED—Socioempathic AI-based Dialogues, Cologne Cobots Lab. The robot head is mounted on a housing incorporating a 10-inch display to show a manual and transcribed dialogue during interactions (see fig. 2). The robot was named ULI, which stands for 'user-language interface', and the name was chosen to be unisex so as to minimize any potential gender biases. The term 'robot' is used throughout this contribution to refer to all elements of the system.

Workshop

In the fall of 2021, we conducted a workshop with employees from the OZEANEUM. In this workshop, we aimed to define visitors' needs by drawing from the employees' experiences in different areas of the museum, such as visitor services, communication and marketing, education, and mediation. The workshop participants, each responsible for different areas of the museum, provided a valuable understanding of visitors' needs and the variations that exist depending on the specific area. Those needs were further refined through subsequent shadowing in various parts of the museum, including at the entrance, cash desk, and information desk, as well as within the exhibitions. This information, combined with resources from the museum (for instance, digital data from museum information brochures and maps), served as the foundation for the prototype's knowledge base.

Figure 2: Images displayed on the tablet. Short interaction manual and example utterances (top). Transcribed text from an example interaction (bottom). Translated from German to English. Source: Ana Müller, Cologne Cobots Lab CC-BY.



The Field Study

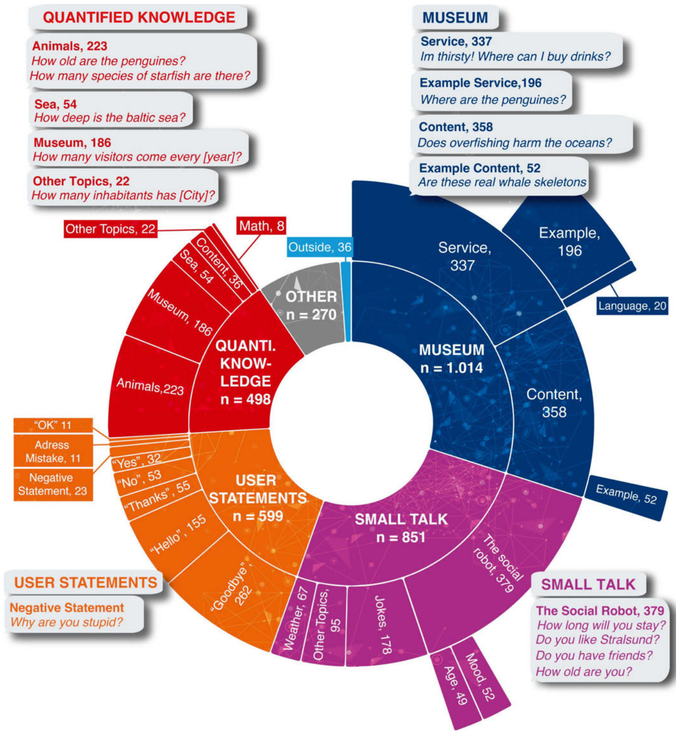
The field study with the dialogue system connected to the Furhat front-end (fig. 1) took place from 10 to 30 March 2022 at the OZEANEUM in Stralsund. The robot was tested at two locations in the exhibition (at the beginning and middle of the tour). During our field study, a total of 11,923 visitors ($M = 568$, $SD = 231.47$ visitors per day) had the opportunity to explore the exhibitions.

We collected data from 3,913 utterances using a logging system: n = 446 utterances were recorded at the beginning of the tour and n = 2,476 were recorded at another location in the centre of the museum; n = 25 interactions were eliminated due to privacy regulations, and n = 620 utterances were eliminated due to unclear speech-to-text processing caused by loud ambient noise or user groups talking to each other while interacting with the system. After being cleaned, the dataset consisted of n = 3,268 valid utterances. At this point, it is important to note that conducting tests in a natural history museum presents unique challenges, since not all visitors are necessarily interested in robotics or AI. On the one hand, this can enhance diverse user feedback and usability testing, which can ultimately lead to improvements in future iterations of the system, but, on the other, may also result in a rather small proportion of visitors interacting with the robot.

The qualitative analysis conducted in this study consisted of a descriptive log analysis to reveal patterns and trends. To categorize the user utterances, a combination of predetermined utterances and inductive coding was used. The predetermined utterances were derived from the system's intended use in the museum and the sample utterances displayed on a tablet (as shown in fig. 2). Inductive coding helped shed light on user behaviour patterns. The transcripts and categories were independently analysed by two authors using open coding so as to identify common user utterance tendencies with a specific focus on the desire for quantifiable information.

Results from Interaction Categories and Codes

Figure 3: Sunburst graphic to illustrate the results from system log analyses. Including categories, codes, and quoted utterances by the users as examples. Source: Ana Müller, Cologne Cobots Lab, CC-BY.



The inner circle of figure 3 shows the categories of user utterances by topic, while the middle circle uses corresponding colours to visually represent the associated codes. The outer circle provides additional information, if necessary, to further enhance understanding of the category and code. To illustrate user behaviour patterns, relevant quotes from user utterances are included on the margin.[2]

Museum

We recorded n = 1,014 utterances that can be categorized within the defined museum category. They are related either to content (n = 358) or service (n = 337). These utterances—and their occurrence in 68.97 per cent of utterances—are closely linked to the primary motivating factors for visiting a museum. Visitors want to learn about the exhibitions and the services offered. Since they come to the museum to obtain information and have a memorable experience, it is thus not surprising that their utterances focus on content and services. We also categorized users asking for help in different languages (n = 20) within this category, since offering help and information in various languages can enhance the service encounter.

Visitors might turn to the robot to find information on exhibitions and services. With the robot readily available, visitors are more likely to ask questions and obtain information, as opposed to searching for information on their own (for instance, such as a sign describing the exhibit or the respective number for the audio guide). Furthermore, the display of the Furhat robot showed two common utterances from this category: 'Where are the penguins?' (service, n = 196) and 'Are these whale skeletons real?' (content, n = 52) (see fig. 2). These questions were defined in advance along with the employees as typical service utterances and included as examples (see section 3.2.). Nevertheless, in the case of example utterances, it is unclear whether visitors asked these questions because they were genuinely interested or were merely testing the robot by asking the example utterances in the manual.

The results, however, show that visitors interacted with the Furhat robot as a source of information- and service-related utterances. Most of the utterances focussed on exhibitions and services, which align with visitors' main reasons for coming to the museum. There was little difference between the utterances addressed by visitors to human employees and those addressed to the robot. It is nevertheless essential to consider diversity in the target audience in the development of the system's knowledge base and training data, as different visitors might have different ways of asking questions. The results highlight the importance of incorporating insider knowledge when developing the system in order to accommodate diverse user behaviour patterns.

2 All quotes from user utterances in this paper have been translated from German to English.

To conclude, visitors at a museum evince a strong interest in learning about the exhibitions and services offered, which is reflected in many utterances about content and requests for information on services such as finding specific exhibitions, food options, and restrooms.

Small Talk

Visitors did, however, also engage in small talk with the robot, including discussions about the weather (n = 65) and jokes (n = 178). Small talk was a frequently occurring category (n = 825) and asking for the weather and telling jokes were example utterances displayed as part of the manual on the tablet (fig. 2). While these categories were not ones that occurred most frequently within the small talk category, the frequency of weather-related utterances is particularly interesting because users often asked for the weather in Stralsund and other cities (for example, their hometowns).

We found that many visitors addressed the robot ULI itself, inquiring about its purpose (n = 379) or its mood and age (n = 52 and n = 49, respectively). For instance, they asked questions such as 'Why are you here?' and 'How long will you stay?' or 'Do you like Stralsund?' and 'Do you miss Cologne?', thus showing a fondness for the anthropomorphic robot. This behaviour aligns with the CASA paradigm, which posits that people tend to attribute humanlike traits, motivations, emotions, or intentions to nonhuman entities. While the system logs only provide transcripts of the interactions and not direct observations, they suggest that users have a significant interest in the supposed personality of the robot, which is also reflected in the communicative level of their interactions. It is unclear, however, to what extent the robot's humanlike appearance contributes to this phenomenon, and/or what contribution the AI's communication design makes and what influence the location, target group, or other influencing factors play or how they converge.

User Statements

To better understand the application of anthropomorphism and to grasp the treatment of social robots in public spaces as either 'machines and servants' or 'entities on their own', a user statements category was defined. The codes were refined to include conversational norms such as greetings and farewells. The analysis showed that many visitors did not follow the same communication norms when speaking to the robot as they would with other humans. The lack of transfer of politeness norms from human-human interactions to human-robot interactions is evident in the statement category analysis and other categories discussed.

Our analysis revealed that only a few users expressed politeness during their interactions with the social robot. For example, users rarely thanked the robot for the information provided. These statements must, however, be considered with respect to all other user utterances, since only 1.59 per cent of the utterances as a whole contained a 'thank you'. This might be because solely user statements recorded by the system were analysed, and statements such as 'hello', 'thank you', and 'goodbye' may not be present in the dataset if they were said when the robot was not processing. Additionally, it is possible that users would have applied politeness norms for human-human interactions and that the lack of such examples in the dataset may be due to a lack of satisfaction with the information provided. This again highlights the challenge of designing knowledge for social robots that meets the expectations of diverse users in public spaces.

Quantified Knowledge

The development of social robots and situated dialogues in public spaces poses several challenges, as discussed in section 2. One of the main difficulties is the need for extensive preliminary work on creating knowledge for robots without a specific operational area. This is due to the limitations of handcrafting training data for diverse user groups. Another challenge arises from varying user expectations and motivations, such as testing the intelligence and capabilities of the robot.

This is also reflected in our results and illustrated in figure 3. According to our analysis, n = 498 user utterances, accounting for 15.24 per cent of the valid utterances, can be categorized as requests for quantified knowledge. This category can be broken down further into various codes, such as the size of animals (n = 223), the size of the OZEANEUM (n = 186) or sea (n = 54), and specific quantifiable information about the exhibit content (n = 36). A small number of utterances (n = 22) were made in connection with quantifiable knowledge outside of the OZEANEUM. A few users asked mathematical questions, such as 'What is 1+1?', which may have been intended as a test of the AI robot.

A more in-depth analysis of the system logs shows that many of the utterances in this category were about the size, speed, or age of large animals or marine life. Users asked, for example, about the growth of different whale species, the swimming speed of penguins, and the life expectancy of polar bears. This pattern may be specific to the OZEANEUM and its target audience, and may not be representative of other use cases or museums with different focuses.

Nevertheless, the frequent occurrence of utterances addressing the size of different whale species can be explained not only by users trying to push the limits of the AI, but also the tremendous presence of the ocean giants. Different whale skeletons of stranded whales could be seen from both locations where the robot was positioned. Additionally, one of the locations for the experiment was at the beginning of the exhibition right next to an escalator with the same length as a blue whale, with a sign on the escalator indicating that it is '... 36 metres long, the same length as a blue whale'. All these factors may thus have affected users' expectations of the system.

At the same time, it is also conceivable that users are influenced by the way information is conveyed in museums in general. After all, brief descriptions of the exhibits frequently provide a summary of respective information. In the case of the OZEANEUM, this is often quantified information about aquariums, living beings, and the cold seas. In other museums, questions about, for instance, the age of artworks may also be common. Visitors also frequently asked about the size, number of floors, and visitor count of the museum, as well as the number of fish in the aquarium and the size of the smallest and largest creatures on display.

Conclusion

Robots provide a convenient and efficient way for visitors to gather information. They can ask their questions and receive immediate answers, which saves time and effort compared with searching for information on their own. One dilemma in connection with such robots is, however, the need for them to have both conversational abilities and comprehensive knowledge of many subjects, while the operator wants to maintain control over the accuracy of the information provided. Based on our experiment, we suggest that the following topics be included in the robots' knowledge base.

In conclusion, the high number of exhibition- and service-related questions addressed to the robot in a museum reflects the visitors' primary goals in coming to the museum, the availability and convenience of information—and trust in technology. The experiments have shown that a service robot should thus not only have knowledge related to its primary task. In this particular use case, which involves providing guidance and service, the robot will inevitably be asked about a wide range of subjects. Our findings indicate that the robot must possess exceptional communication skills and a good understanding of various topics.

Additionally, the robot should have information about its immediate environment, or at least give users suggestions regarding where to find this information. These requirements for the knowledge base of social robots in museums can, however, vary depending on the location of the museum, and we assess that the need will be higher in tourist areas. The latter also applies to the region where the OZEANEUM is located, which is why some utterances were asked about nearby tourist attractions, restaurants, et cetera.

The robot should be able to adequately respond to user statements. Besides 'hello', 'thank you', and 'goodbye', it should also be able to respond to 'Yes', 'No', 'Ok', and similar utterances, and be able to consider these within the context of the conversation. For a natural conversation, it is thus essential to establish mixed-initiative and not solely master question answering initiated by the users' utterances.

In this context, the ability to engage in small talk is also important. Not all users see the system as a source of knowledge or services. They might simply want to feel like they are having a conversation. The ability to engage in small talk is thus a crucial aspect in realizing an entertainment value. The robot itself is also of interest to users, and creating a robot personality might help satisfy users' needs for anthropomorphism.

As our results show, establishing what we call 'quantified knowledge' is essential as well. The museum and its contents are not sufficient in this respect. Whether based on the real interest of users, for entertainment purposes, or as a test of the AI, users expect detailed information about the quantifiable content of the museum or exhibit. This includes all the information on the size and age of the items displayed and on the museum itself. Professional information management and effective digital database systems are therefore advantageous for museums.

Moreover, the operator desires control over the accuracy of the information provided by robots. These areas should thus be prioritized when incorporating knowledge into the robots' programming. We highly recommend gathering the information for the topics mentioned through workshops with the staff, reviewing frequently asked questions, and visiting the operational site through the eyes of a regular visitor. Nonetheless, the manual creation of training data for AI systems is a labour-intensive task that is also influenced by the developers' demographics, such as age, education, or profession. As a result, the knowledge base may only work well for the specific target group to which the developer belongs. Within the research project SKILLED, we are therefore currently developing a method to improve the system's knowledge while ensuring quality assurance by using a distributed system. In future, the system will feature a knowledge base tailored to a specific use case, a powerful language-based question-answer model, and, potentially, links to other AI systems like ChatGPT. It is, however, also crucial to also consider ethical aspects, like preventing the dissemination of inappropriate jokes and striking a balance between brief and comprehensive answers, some of which may not always be suitable for public spaces and changing circumstances.

References

Cantucci, Filippo/ Falcone, Rino (2022). Autonomous Critical Help by a Robotic Assistant in the Field of Cultural Heritage: A New Challenge for Evolving Human-Robot Interaction. Multimodal Technologies and Interaction 6 (8), 69. https://doi.org/10.3390/mti6080069 (all URLs here accessed in August 2023).

Epley, Nicholas/Waytz, Adam/ Cacioppo, John T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism. Psychological Review 114 (4), 864–86. https://doi.org/10.1037/0033-295X.114.4.864.

Foster, Mary Ellen (2019). Natural Language Generation for Social Robotics: Opportunities and Challenges. Philosophical Transactions of the Royal Society B: Biological Sciences 374 (1771), 20180027. https://doi.org/10.1098/rstb.2018.0027.

Gao, Jianfeng/Galley, Michel/Li, Lihong (2018). Neural Approaches to Conversational AI. Proceedings of ACL 2018, Tutorial Abstracts, 2–7. https://doi.org/10.18653/v1/P18-5002.

Ivanov, Stanislav/Webster, Craig/Berezina, Katerina (2017). Adoption of Robots and Service Automation by Tourism and Hospitality Companies. In: Revista Turismo & Desenvolvimento 27 (28), 1501–17. Available online at https://ssrn.com/abstract=2964308.

Lee, Kwan Min/Peng, Wei/Jin, Seung-A et al. (2006). Can Robots Manifest Personality? An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction, Journal of Communication 56 (4), 754–72. https://doi.org/10.1111/j.1460-2466.2006.00318.x.

LivePerson Inc (Oct. 2022). Intent Manager, Best Practices, LivePerson Developer Center. Available online at https://developers.liveperson.com/intent-manager-best-practices.html.

McTear, Michael/Callejas, Zoraida/Griol, David (2016). The Conversational Interface: Talking to Smart Devices. Cham, CH, Springer https://doi.org/10.1007/978-3-319-32967-3_4.

Minstrom, Michael/Sumartojo/Shanti/Kulić, Dana et al. (2022). Robots in Public Spaces: Implications for Policy Design, Policy Design and Practice 5 (2), 123–39. https://doi.org/10.1080/25741292.2021.1905342.

Nass, Clifford/ Brave, Scott (2005). Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship. Cambridge, MA, The MIT Press.

Reeves, Byron/ Nass, Clifford (2006). The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places. Stanford, CA, CSLI Publications.

Sabanovic, Selma/Michalowski, Marek P./Simmons, Reid (2006). Robots in the Wild: Observing Human-Robot Social Interaction Outside the Lab. 9th IEEE International Workshop on Advanced Motion Control, 596–601. https://doi.org/10.1109/amc.2006.1631758.

Skantze, Gabriel/Al Moubayed, Samer/Gustafson, Joakim et al. (2012). Furhat at Robotville: A Robot Head Harvesting the Thoughts of the Public through Multi-Party Dialogue. Proceedings of the Workshop on Real-time Conversation with Virtual Agents IVA-RCVA. Available online at https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=cfbaf70d801b8237dd268b83ac3d524d7c49c9c8.

Smithsonian Institution (2018). Smithsonian Launches Pilot Program of 'Pepper' Robots. Softbank Robotics Donates Robots as Part of Significant Gift. Available online at https://www.si.edu/newsdesk/releases/smithsonian-launches-pilot-program-pepper-robots.

Tian, Leimin/Carreno-Medrano, Pamela/Sumartojo, Shanti et al. (2020). User Expectations of Robots in Public Spaces: A Co-design Methodology. Proceedings 12th International Conference, ICSR 2020, Golden, CO, USA, November 14–18, 2020, 259–70. https://doi.org/10.1007/978-3-030-62056-1_22.

Villaespesa, Elena (2021). List of Artificial Intelligence (AI) Initiatives in Museums. Available online at https://www.artsmetrics.com/en/list-of-artificial-intelligence-ai-initiatives-in-museums/.

Willeke, Thomas/Kunz, Clay/Nourbakhsh, Illah (2001). The History of the Mobot Museum Robot Series: An Evolutionary Study. Proceedings of the 14th International Florida Artificial Intelligence Research Society Conference (FLAIRS '01), 514–18. Available online at https://www.ri.cmu.edu/pub_files/pub3/willeke_thomas_2001_1/willeke_thomas_2001_1.pdf.