Adaptive AI Museum Guides: Enhancing User Experience with Furhat Robots

MATTEO DEL PRATO
matteodp@kth.se

ALESSANDRO AMANDONICO
aleama@kth.se

MALIN BRILON
brilon@kth.se

STIJN TEEKENS
teekens@kth.se

February 21, 2025

Abstract

This project focussed on the implementation of a virtual Furhat robot as an adaptive personal museum guide in a virtual museum environment. Two personalities were designed to study the effects of variations in speech in- and output modalities and interaction style on the engagement and perceived education of visitors. Furhat A portrayed a calm and reflective guide, whereas Furhat B portrayed a more lively and "cool" persona. 7 participants of varying age interacted with both robots in a controlled virtual museum environment and provided feedback through a questionnaire containing UEQ+, GODSPEED, and self-perceived learning assessment questions. The results indicated a generational difference in preference, as participants of 30 years or younger favoured Furhat B, whereas participants older than 30 appreciated the calmer Furhat A more. Overall, Furhat B outperformed Furhat A in terms of efficiency, engagement, and perceived educational value. Limitations of this study concern the small sample size, the variability in participant familiarity with AI, and some occasional technical issues. Future research could entail a more enhanced personalisation of the Furhat personality, more elaborate prototype testing, and improving the accessibility for diverse audiences. Overall this research contributes to the design of interactive, and educational experiences in virtual museum contexts.

Contents

1 Introduction

Not all museums are evenly successful at recovering from the pandemic [1] and it can be a real struggle to reattract returning visitors, or to successfully engage certain groups of visitors, such as teenagers [2] [1]. Furthermore, currently there are globally approximately 2.2 billion people that suffer from a near or distance vision impairment [3]. Museums are often not equipped to provide the same kind of engaging experience that non-visually impaired visitors can enjoy, and therefore new tools have to be utilized to be able to let a museum visit be an educational, enjoyable, and inclusive experience [4]. To attract and retain visitors and give them a new experience a social robot could be employed, one such as Furhat.

1.1 Background

Several studies provide valuable insights into the design of voice-driven systems and robots for enhancing museum experiences, particularly in terms of user engagement, personalization, accessibility, and content delivery, offering a comprehensive framework for our project goal and structure.

Barth et al. [5] investigate the type of content that voice-based AI conversational systems should include to meet visitors' expectations in museums. Here, much stress is put on the importance of not only providing information about exhibits but also explaining the meanings and intentions behind the artworks. This aligns with the goals of our project, which aims to provide informative responses about paintings in a virtual museum. By addressing common questions and using a standardized approach to handle various user queries, we can ensure that the Furhat robot delivers a relevant and educational experience. The study's evaluation methodology, which uses user satisfaction and interaction quality to assess the effectiveness of the system, will be useful in evaluating the impact of different Furhat personalities and their responses.

Niculescu et al. [6] highlight that the attractiveness and engagement of social robots are influenced by factors like voice pitch, humour, and empathy. Their study suggests that higher-pitched voices can be more engaging and appealing, and incorporating humour can enhance the overall user experience. For our project, varying the voice pitch according to different robot personalities—such as a youthful, energetic voice for younger visitors or a calm, soothing voice for older ones—could help make interactions more engaging. Additionally, integrating humour and empathy will create a more relatable and enjoyable experience for users, particularly during personalised museum tours. These elements will be critical in ensuring the Furhat robot is engaging across different user demographics.

In a similar vein, Gaia et al. [2] explore the design of chatbots for museum engagement, specifically targeting teenagers. Their research shows that interactive, informative, and adaptable systems are crucial for attracting and retaining young audiences. This research suggests that the Furhat robot should not only respond to predefined questions but also engage visitors in dynamic, conversational interactions based on their interests. They also advocate for incorporating gamification features, such as tracking points or solving mysteries, to make the museum experience more interactive. These findings could help us design a more engaging, flexible system that encourages active participation among visitors, especially younger users.

Ensuring accessibility is another essential consideration for designing inclusive museum experiences. Adcock [7] discusses the development of a voice-driven smart guide for blind and partially sighted (BPS) visitors. The guide allows users to navigate exhibits through voice commands and provides clear, descriptive audio of the exhibits. This approach will be particularly useful for our Furhat robot, as it must ensure that detailed, audio-based descriptions are accessible to all users, including those with visual impairments. Adcock's study also highlights the importance of privacy protection in voice interactions, a concern that Kulathunga [8] emphasizes as crucial in building user trust.

While many of the studies focus on interaction and accessibility, Peponis et al. [9] focus on the role of spatial design in shaping visitor behaviour and engagement in physical museum spaces. Their research

suggests that the strategic placement of exhibits in open-plan settings can guide visitors through a curated journey, improving their learning and engagement. While our project is based in a virtual environment, these principles are still applicable. For example, we can apply spatial design concepts in the virtual space to create a logical flow between exhibits, subtly guiding users through different sections of the museum while maintaining a sense of openness and discovery.

Kulathunga [8] also provides insights into customizing robot personalities for different educational contexts. Their research underscores that user engagement is influenced by the robot's tone, pace, and emotional expression. This is particularly relevant for our project, where the ability to adjust the Furhat robot's personality based on user preferences—whether they prefer a more direct, informational style or a narrative-driven approach—will be essential for maximizing engagement. Furthermore, balancing the robot's pacing to avoid frustrating or boring users will be critical.

Overall, we aim to create a museum experience that is both educational and enjoyable for a wide range of visitors by designing an interaction that is not only engaging but also personalized and accessible.

1.2 Modalities

The choice of modalities in our study —speech input/recognition and speech output/synthesis— was motivated by the goal of providing an intuitive and accessible means of interaction for all museum visitors, including those with visual impairments. Speech, as a natural form of communication, offers a direct and personal way for users to engage with the Furhat robots, ensuring an inclusive and immersive experience. Given the complex and multi-layered nature of museum environments, speech recognition allows users to navigate through the exhibits and interact with the robot, while speech synthesis enables the robot to respond in an entertaining and engaging way, influencing the learning experience.

Speech as a modality has been widely used in human-robot interaction (HRI) studies because it aligns with the natural human communication process. According to related work, this modality allows for more dynamic and flexible interactions compared to traditional button-based or touchscreen interfaces [6]. Furthermore, voice-driven systems are especially valuable in environments where users may be unable to rely on visual cues, such as in the case of blind or partially sighted museum visitors. Adcock [7] highlights the significance of voice interaction in creating accessible museum experiences by providing clear, descriptive audio. Our decision to implement this modality aligns with their findings, ensuring that all visitors, regardless of their abilities, can benefit from the assistant's guidance. In our implementation, we utilized two different Furhat robots with distinct personalities to examine the impact of personality-based variation in communication style. The robots were programmed with different speech profiles: Furhat A, with a calm and reflective persona, and Furhat B, with a lively and energetic tone. These personas were integrated into the system using pre-configured scripts that alter speech pitch, pace, and language style, based on the desired character traits. As demonstrated in related work, varying speech characteristics, such as voice pitch and tone, can influence user engagement and comfort [6]. For example, a calm voice may appeal to older visitors, while a more energetic voice could resonate with younger, more tech-savvy individuals [2].

We also compared these two different implementations by evaluating user perception and experience. Previous studies in HRI suggest that varying the robot's personality can affect how users interact with the system and how engaged they feel during interactions. Syrdal et al. [5] found that robot personalities that align with user expectations and communication preferences result in more effective and enjoyable interactions. For this reason, our study compares how participants respond to these two distinct personas—Furhat A and Furhat B—in terms of user preference, perceived interaction quality, and educational value.

In terms of the implementation of these modalities, we used the Furhat robot's built-in speech synthesis

and recognition software, coupled with natural language processing (NLP) to handle user queries. The system was designed to recognize both specific and vague requests from users, processing information based on a structured dataset of museum content. By comparing the interaction between two distinct robot personalities, we aim to investigate which persona fosters better engagement, retention of information, and personalization of the museum experience, particularly for different age groups and user preferences.

Additionally, the choice to compare modalities in this study was motivated by the findings of previous research that underscore the importance of user-centered design in technology adoption. For instance, Nomura et al. [9] and Niculescu et al. [6] emphasize that voice pitch, humour, and emotional cues in robot interactions can enhance user satisfaction and engagement, particularly in settings like museums where visitor experience is paramount. Our work builds on these ideas by experimenting with personalized, voice-driven interactions through the use of distinct robot personalities.

The comparison of modalities in this study serves to explore the effectiveness of personalized robot personalities in facilitating a more enjoyable and educational museum experience. The combination of user feedback from surveys (e.g., the UEQ+ and GODSPEED scales) and data analysis (e.g., interaction frequency and time spent) will allow us to assess how varying robot characteristics influence the quality of the interaction and users' perceived value of the educational content provided by the robot.

Overall, the selected modalities are designed to provide an engaging, personalized, and accessible experience for a wide range of users. By comparing these modalities, we aim to offer insights into how personalized voice-driven systems can enhance the museum experience, not only for the general public but also for specific visitor groups, such as younger audiences and individuals with visual impairments.

1.3 Hypotheses

To achieve this goal, several hypotheses have been proposed:

1. H1: Participants older than 30 will prefer the calm, older Furhat A. Whereas participants of 30 years or younger will prefer the more active, younger Furhat B (*User Preference*)

2. H2: There will be a difference in naturalness and ease of interaction scores between Furhat A and Furhat B (*interaction quality*)

3. H3: Participants will find the personalized museum path provided by the Furhat robots to be useful (*personalized experience*)

4. H4: The effectiveness of the personalization (e.g., suitability of suggested paths) will vary based on the respondents' age and interests (*personalized experience*)

5. H5: There will be a difference in the perceived educational value of interactions with Furhat A and Furhat B, with participants rating one of the two as more informative based on personality and engagement levels. (*educational value*)

2 Method

This study aims to evaluate user preferences, quality of interaction, and perceived educational value of two Furhat robots - Furhat A, Jenna characterized as a quiet, older persona, and Furhat B, Bill characterized as an active, younger persona - in a virtual museum environment. The methodology includes building of the prototype, participant recruitment, preparation of the experimental environment, interaction sessions, and data collection to ensure a rigorous and systematic approach to investigating the hypotheses outlined.

2.1 Building the Prototype

The prototype for the museum assistant was developed starting from the Furhat lab material to offer an engaging interface that would enhance the user's exploration of a virtual museum. Its design focuses on providing users with comprehensive information about artists, artworks, and the museum's layout while ensuring seamless navigation and interactive engagement.

The assistant leverages a structured dataset that organizes artists, artworks, and museum areas into distinct categories. This dataset serves as the backbone for all interactions, allowing the assistant to respond accurately to user queries. The structure of the dataset is shown in Table 1.

Category	**Details**
Artists	30 (Name, Artworks)
Artworks	40 (ID, Title, Author, Year of Production, Museum Wing)
Museum Wings	4 (Cubism, Renaissance, Romantic, and Surrealism wings)

Table 1: Dataset Structure

The system dynamically recognizes various entities, such as artwork titles, artist names, and numerical identifiers for artworks, enabling it to handle both precise and vague user inputs. For instance, when a user asks, "Who painted artwork number 5?" the assistant cross-references the ID with its dataset to deliver the correct response.

One of the key features of the prototype is its ability to provide detailed information about the museum's collection. Users can ask about the list of artists featured in the museum, inquire about specific artists and their works, or request the locations of particular pieces. The assistant can also provide historical details, such as the creation year of an artwork, and confirm whether specific artists or pieces are part of the museum's collection.

The virtual assistant also incorporates navigation capabilities, offering detailed insights into the museum's structure. It explains the different areas or wings, such as the "Cubism Wing" or "Renaissance Gallery," and lists the artworks displayed in each section. Additionally, it tracks the user's progress through the museum, identifying which areas have already been visited and which remain unexplored. This feature ensures that users can efficiently organize their virtual tour and make the most of their experience.

To further enhance interactivity, the assistant supports guided tours, allowing users to start and end their visits seamlessly. During the tour, the assistant adapts its responses based on the user's preferences, providing curated information about artworks, their creators, and their locations. To do so, Furhat employs the use of OpenAI APIs to use an LLM that has different prompts based on the personality of the virtual assistant used. The prompts used by the LLMs are the following:

- **Furhat A (calm, old lady):** "Respond in the style of a wise and passionate old character named Jenna. Speak warmly and thoughtfully, with a deep appreciation for art, culture, and life. Use poetic and descriptive language, weaving in personal anecdotes or philosophical musings. Be encouraging and reflective in your tone."
- **Furhat B (active, young boy):** "Respond in the style of an enthusiastic and modern young character named Bill. Use energetic, casual, and witty language, with a touch of humor and modern slang. Be engaging, friendly, and relatable. Feel free to make jokes or refer to pop culture to keep the conversation lively."

In constructing this prototype, intent recognition and natural language processing played a central role. The assistant identifies the purpose behind queries, whether it is a request for artist details, an inquiry about

an artwork's location, or a navigation-related question, and generates contextually relevant responses. By combining precise information retrieval with conversational engagement, the system creates an intuitive and user-friendly experience. The interaction flow for the engagement of the Furhat with the visitor had been carefully designed in Miro [10]. A selection of the board is illustrated in 1. During the design process it was taken into account the fact that the Furhat not only had to answer informational questions (for which it had to utilize its large language model functionality) but also questions specifically related to the virtual museum the visitor was interacting with. For this we could apply the knowledge that was gained during the lab session, for instance about how to account for similar phrasings of one particular question.

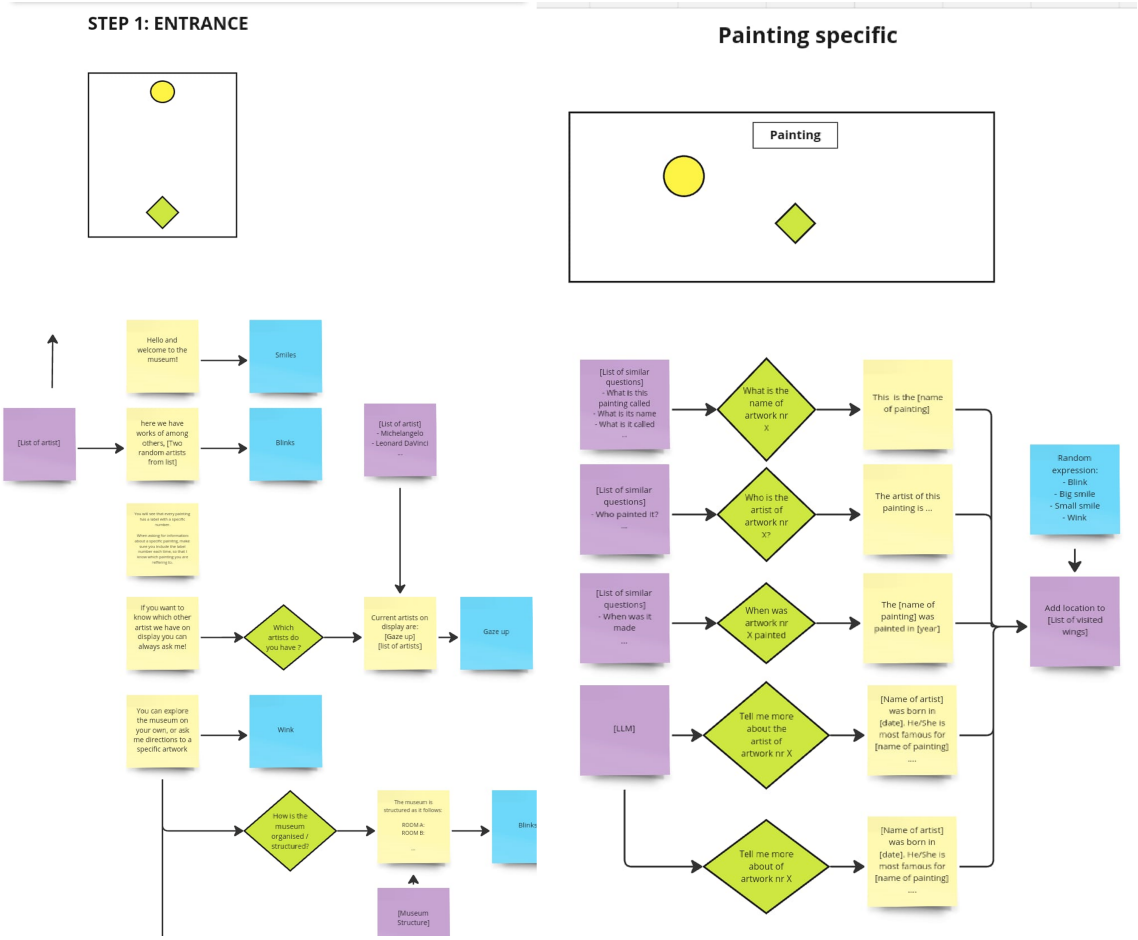

Figure 1: Part of the interaction flow design. Entrance: start of the interaction flow between the Furhat and a visitor. Painting Specific: visitor ask the Furhat about a certain painting

2.2 Participant Recruitment

The participants were recruited with convenience sampling to represent different age groups and levels of familiarity with technical devices. A pre-test screening collected demographic information such as age, gender, education level, and familiarity with AI social service robots. The survey also assessed participants' attitudes towards robots using the Negative Attitudes Towards Robots Scale (NARS) [11, 12] and evaluated the participants' familiarity and experience with robots [13].

2.3 Experiment Environment

The experiment was conducted in a quiet, controlled environment to minimize distractions. The setup included two Furhat robots and an interactive PowerPoint presentation that simulated a virtual museum gallery. Technical checks were carried out before each session to ensure that all functions, including speech recognition, speech synthesis, and navigation, were operational. The speech recognition system was calibrated to account for different accents and voice variations.

2.4 Interaction Sessions

Participants took part in two consecutive interaction sessions, each involving one of the Furhat robots. The sessions were designed to evaluate the hypotheses related to user preference, interaction quality, personalization, and educational value.

Structure of the sessions:

1. **Session 1: Interaction with Furhat A**
    - Participants were greeted by Furhat A and introduced to the virtual museum.
    - They navigated through the gallery and asked Furhat A questions about the artworks.
    - Furhat A provided answers and suggested a personalized museum path based on the participants' stated preferences.
    - Participants completed a survey after the session to record their experience.
2. **Break:** A 5-minute break was taken to counteract fatigue.
3. **Session 2: Interaction with Furhat B**
    - The same interaction procedure was followed as in Session 1, but participants interacted with Furhat B.
    - A second survey identical to the first one was conducted after the session.

To control for order effects, the order of the interaction sessions (i.e., starting with Furhat A or Furhat B) was randomized between participants.

2.5 Data Collection

A survey was conducted prior to the user test and following each interaction with the Furhats. The complete survey is available in .2, while .3 provides a detailed justification and outlines how each question corresponds to specific hypotheses.

2.5.1 Data before the test

Demographic data, the participants familiarity with AI robots [13] and the participants initial attitudes towards robots [14, 11] were collected as part of a pre-test survey.

2.5.2 Post-Interaction Surveys

Following each session, participants completed customized surveys using established instruments:

- **UEQ+:** Measured interaction quality, including efficiency, usefulness, intuitive use, response behaviour, response quality, and comprehensibility [15].
- **GODSPEED:** Assessed anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of the robots [16].
- **Self-Perceived Learning Scale:** Evaluated the perceived educational value of interactions with the Furhat robots [17].

2.5.3 Interaction Metrics

During the sessions, interaction data such as the time spent interacting, the frequency and type of questions asked, and the participation of the participants were recorded. Video and audio recordings facilitated detailed observational analysis, while technical logs documented any system errors or performance issues.

2.6 Data Analysis

The quantitative data from the surveys were analyzed using statistical methods to compare the interaction quality, personalization effectiveness, and educational value of the two Furhat robots. Correlation analyses examined the relationships between participants' demographic characteristics (e.g., age, familiarity with technology) and their responses.

2.7 Ethical Considerations

Participants gave their informed consent before participating in the study and data confidentiality was ensured. The video and audio recordings were stored securely and used for research purposes only. The consent form can be found in .4. Ethical guidelines, including compliance with GDPR, were strictly followed throughout the study.

By combining qualitative and quantitative methods, this study provides a comprehensive evaluation of the user experience with Furhat A and Furhat B, addressing the hypotheses and providing information on possible improvements in the design of human-robot interaction.

2.8 Demo video

A short demonstration of the interaction can be found here [18].

3 Results and Analysis

3.1 Sample demographics

The sample of participants in the test consisted of 7 individuals. Among these participants, the majority were female, with only two males included (see Figure 11 in .5). The ages of the participants were quite varied, ranging from teenagers to individuals between 22 and 25 years old, as well as two participants aged 55 and 59, respectively (see Figure 10 in .5). Most participants had completed either a Bachelor's degree or a graduate degree at the time of the study, while only one participant had recently graduated from high school (see Figure 12 in .5).

3.2 Perception of Robots

Participants' familiarity, experience and attitude towards AI robots were collected as part of a pre-test survey. This consisted of two tests, familiarity and experience with robots [13] to measure consumer trust in interaction with artificially intelligent social robots (see Table 2) and NARS [16][11] to detect negative attitudes towards robots (see Table 3), using in both cases a 5-point Likert scale (see Table 4)

3.2.1 Familiarity and experience with Robots

As shown in Figure 2, across all participants, the average familiarity levels suggest a moderate understanding of AI social service robots.

Younger participants (< 30, Cluster 1) generally rate their familiarity with AI social service robots higher, with scores frequently reaching four or above across the provided familiarity measures; in contrast, older participants (> 30, Cluster 2) tend to score lower in familiarity, reflecting a trend where age and exposure intersect to influence familiarity.

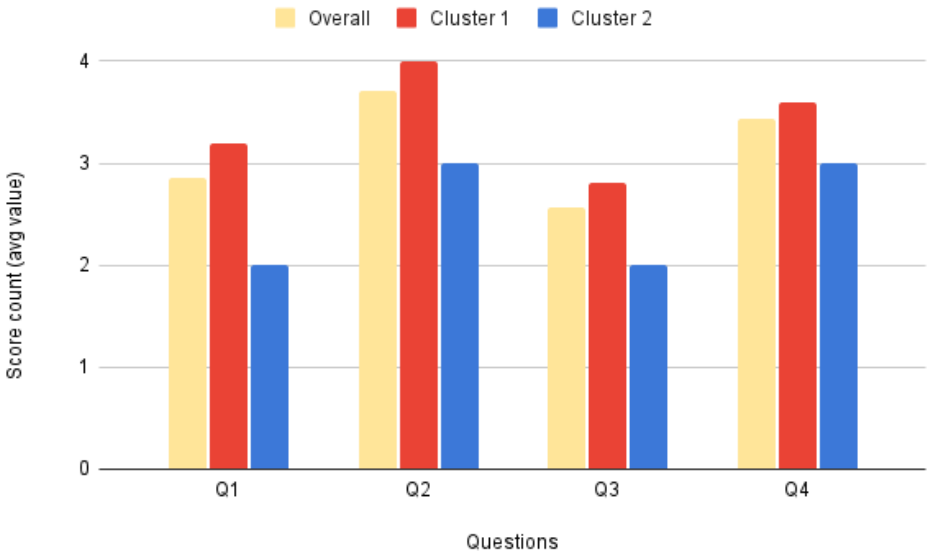

Figure 2: Participants' average score distribution across Familiarity and experience with Robots" questionnaire

3.2.2 NARS:Measure Attitude towards Robots

Figure 3 suggests an overall neutral attitude towards robots, with an average value of 2.62 across all participants.

Interestingly, older participants demonstrate a slightly higher level of comfort and perceived adaptability towards robots compared to younger participants. Specifically, older individuals are more inclined to believe that they could form friendships with robots if the robots had emotions, which contrasts with the views of younger participants.

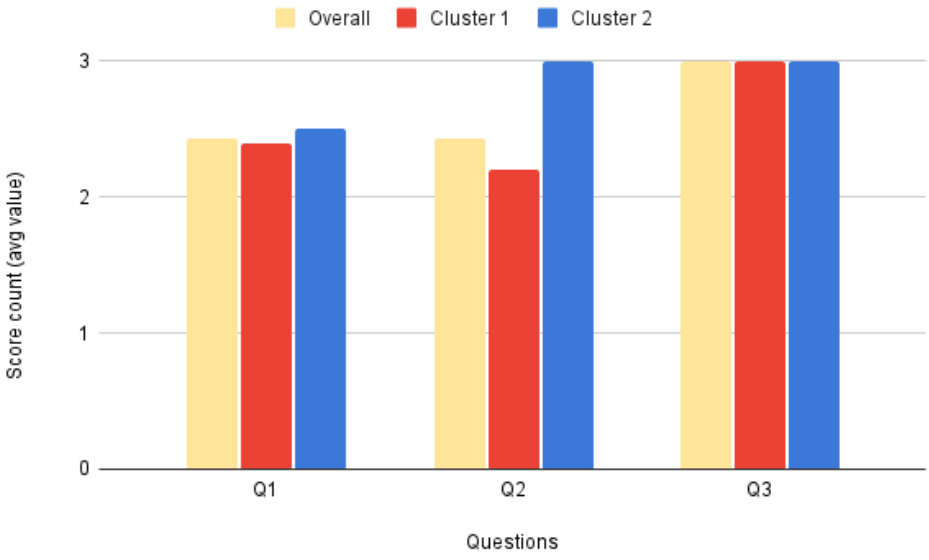

Figure 3: Participants' average score distribution across NARS questionnaire

3.3 Post-Interaction Surveys

3.3.1 UEQ+

As shown in Figure 5, Furhat B generally outperforms Furhat A (see Figure 4 in several areas, with higher scores in efficiency, usefulness and intuitive use, suggesting that participants found it more effective and easier to engage with. Comprehensibility, while consistent across both systems, shows Furhat B slightly ahead, indicating better clarity in responses. Overall, Furhat B is rated more positively, particularly in its practical usability and emotional engagement, though significant variability in individual experiences remains evident.

When examining Furhat A across clusters, younger participants in Cluster 1 generally provide higher scores across most metrics (see Figure 13, reflecting a more favorable impression compared to older participants in Cluster 2. This trend is particularly pronounced in intuitive use, where younger participants report significantly better experiences, and in efficiency, where they also provide higher ratings despite notable variability. In contrast, Cluster 2 (see Figure 14 reports lower scores in usefulness and response behavior, indicating less perceived effectiveness and engagement. Younger participants thus appear to have a more positive interaction with Furhat A, while older participants view it as less usable and less effective.

A similar cluster-based comparison for Furhat B highlights that younger participants in Cluster 1 again provide higher scores (see Figure 15), particularly in intuitive use and usefulness, suggesting greater ease of interaction and utility. Older participants in Cluster 2 report lower scores (see Figure 16, particularly in response behavior and response quality, where averages are closer to neutral or slightly negative. Overall, Furhat B is perceived more positively by younger participants, while older participants provide less favorable assessments, particularly regarding its engagement and utility.

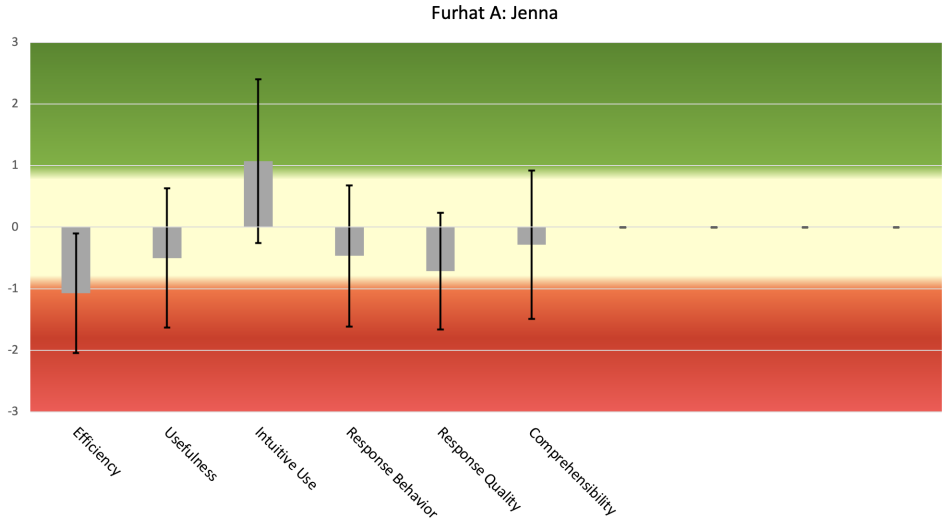

Figure 4: Furhat A UEQ+ overall results

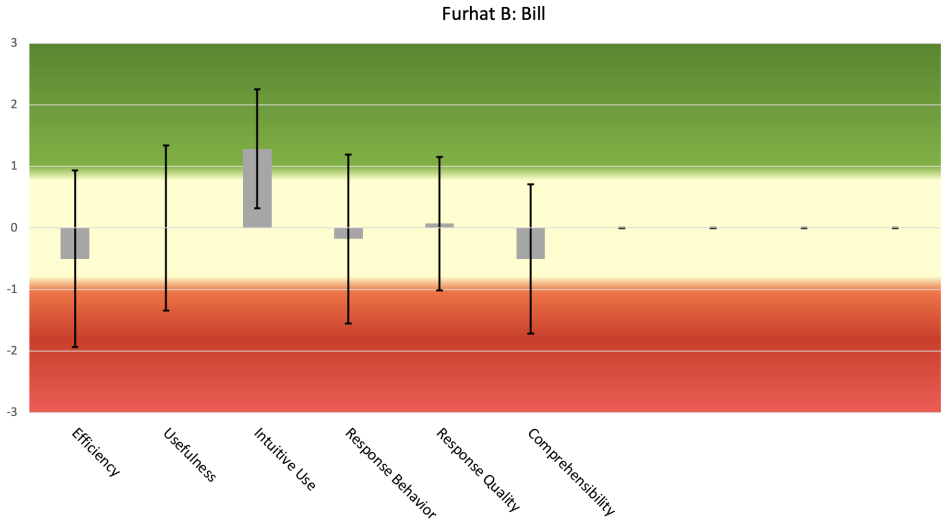

Figure 5: Furhat B UEQ+ overall results

3.3.2 GODSPEED

The analysis of Furhat A and Furhat B's overall performances across all participants (see Figure 17) reveals that Furhat A generally scores higher in anthropomorphism, animacy, and likeability, indicating that participants found it more engaging, lifelike, and likeable. Furhat B, however, achieves higher scores in perceived safety and perceived intelligence, suggesting participants view it as safer and more intelligent. Both robots show variability across the attributes, reflecting differences in user preferences and experiences.

Younger participants This indicates that younger participants found Furhat A more engaging and likable than the older cluster (see Figure 6). On the other hand, older participants valued perceived intelligence and safety attributes more in Furhat A if compared to the other cluster.

For Furhat B, younger participants also provide higher scores in anthropomorphism and animacy, showing a preference for its engaging and lifelike characteristics. However, the older cluster rates Furhat B higher in perceived safety and perceived intelligence, aligning with their preference for these attributes.

Overall, Furhat A outperforms Furhat B in anthropomorphism and likeability in both clusters, while Furhat B scores higher in perceived safety and perceived intelligence regardless of the cluster. Younger participants strongly prefer Furhat A's engaging and likeable qualities, while older participants show a greater appreciation for Furhat B's safety and intelligence, suggesting that Furhat A excels in attributes associated with emotional engagement, whereas Furhat B is preferred for functional and safety-related features, with generational differences playing a significant role in these evaluations.

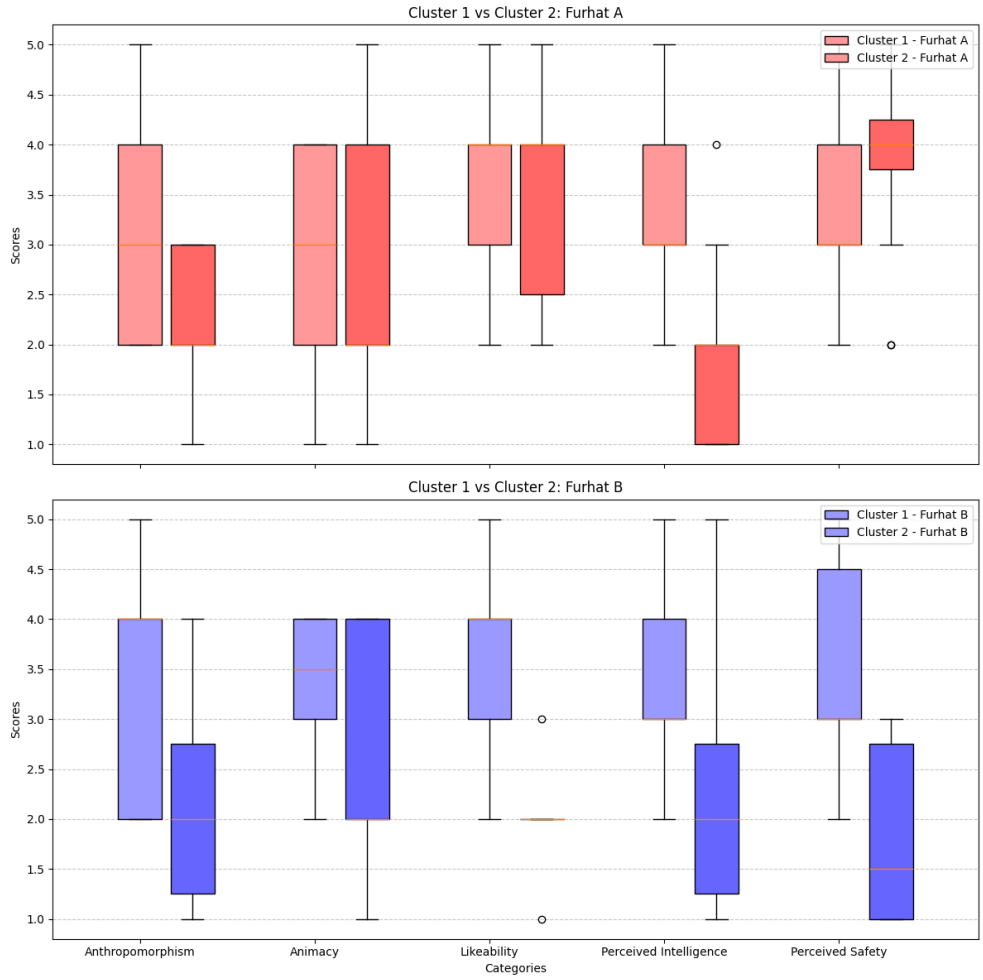

Figure 6: Comparison of results across Clusters for Furhat A(top image) and Furhat B(bottom image)

3.3.3 Self Perceived Learning

According to scores from Self-Perceived Learning assessment, Furhat B consistently scores higher in enhancing understanding and knowledge compared to Furhat A, especially in the categories of terminology, concepts, and overall knowledge.

When focusing on perceptions of Furhat A by participant clusters (see Figure 7), younger participants in Cluster 1 generally rated the robot higher than older participants in Cluster 2. Cluster 1 participants showed better engagement and understanding, particularly in overall enhanced knowledge and understanding, where their scores are noticeably higher. This indicates that Furhat A resonates more effectively with younger individuals, possibly due to its interaction style or content delivery being better aligned with their learning preferences.

For Furhat B (see Figure 8), a similar trend is observed where younger participants in Cluster 1 tend to provide higher scores than those in Cluster 2. However, the difference between the clusters is less pronounced for Furhat B compared to Furhat A, indicating that Furhat B has a more uniform impact across age groups. In particular, Furhat B scores for participants in Cluster 2 are higher than those of Furhat A, suggesting that Furhat B is better received by the older demographic.

Overall, Furhat B consistently outperforms Furhat A in both Cluster 1 and Cluster 2, with the gap being wider in Cluster 2. This highlights Furhat B's superior ability to cater to diverse age groups, while Furhat A seems to be more appealing to younger participants. While Furhat A has strengths with younger participants, Furhat B demonstrates a broader and more consistent appeal, making it the more effective system overall.

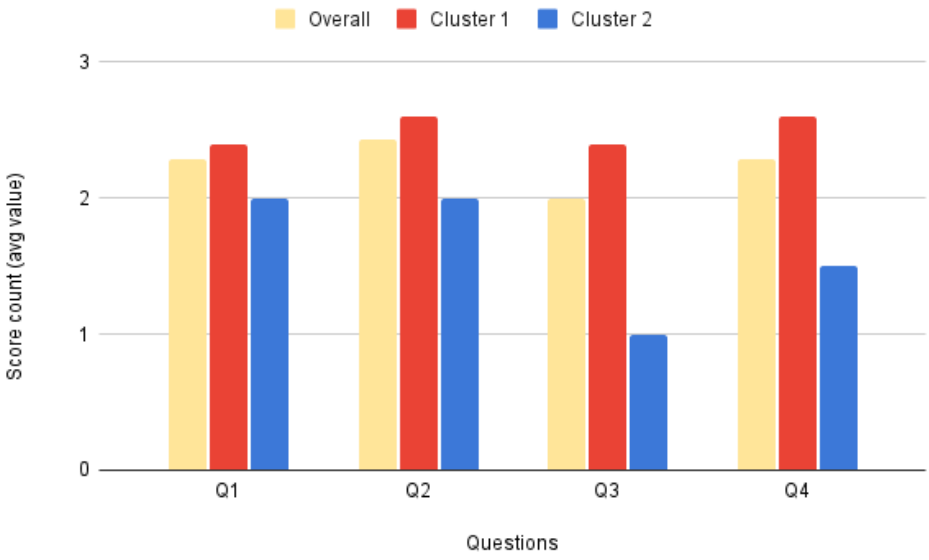

Figure 7: Furhat A: comparison of Self-Perceived Learning scores

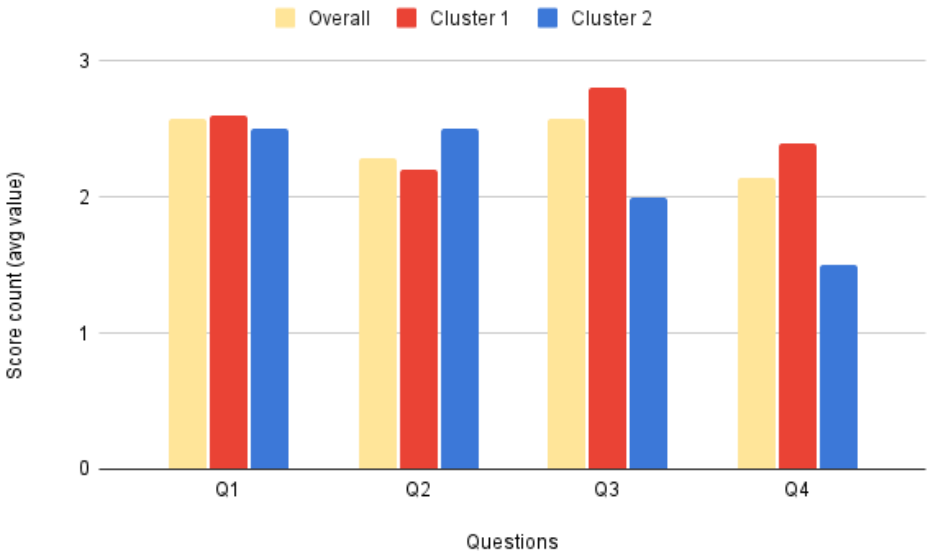

Figure 8: Furhat B: comparison of Self-Perceived Learning scores

4 Discussion

The user test results showed that participants generally had a neutral-to-slightly-positive perception of robots. Older participants reported feeling more comfortable with robots, which influenced their engagement levels. Younger participants were less inclined to form emotional connections with the robots compared to older participants.

Furhat B was preferred overall, scoring higher in efficiency, usefulness, and intuitive use, especially among younger participants. In contrast, Furhat A was less engaging but performed better among older participants due to its calm and reflective personality. Younger participants preferred Furhat B's energetic and modern style, while older participants resonated more with Furhat A. However, older participants still rated Furhat B higher in terms of usability.

Furhat B excelled in response behavior and response quality, delivering clearer and more engaging interactions. Furhat A maintained consistent comprehensibility but was rated lower for practical usability. Regarding educational value, Furhat B was perceived as more effective by younger participants, while older participants rated both robots similarly in this aspect.

Discussing our hypotheses H1: "Participants older than 30 will prefer the calm, older Furhat A and participants of 30 years or younger will prefer the more active, younger Furhat B" it was partially supported. Younger participants clearly preferred Furhat B, but older participants showed only a slight preference for Furhat A, which was not as pronounced.

H2: "Participants older than 30 will prefer the calm, older Furhat A. Whereas participants of 30 years or younger will prefer the more active, younger Furhat B" could be verifiyed the UEQ+ analysis reveals that Furhat B performs better in response behaviour and quality. The GODSPEED confirms this in animacy and anthropomorphism.

The H3: "Participants will find the personalized museum path provided by the Furhat robots to be useful" could not be tested due to time constraints. The personalized museum path was not implemented, but this remains an opportunity for future research.

In H4 we stated that ”The effectiveness of the personalization will vary based on the respondents’ age and interests”, nevertheless this hypothesis was not verified. While the effectiveness of the interactions varied among participants, there were no meaningful patterns related to age or interests.

We could verify H5: ”There will be a difference in the perceived educational value of interactions with Furhat A and Furhat B, with participants rating one of the two as more informative based on personality and engagement levels” because from the “self-perceived learning” questionnaire, Furhat B was found to be more informative than Furhat A.

The speech-based interaction modality proved effective, particularly for participants with limited reliance on visual cues, enhancing the accessibility of the experience. While the use of distinct robot personalities succeeded in addressing a wide range of demographics, the results also highlighted generational differences in interaction preferences. However, the study faced several limitations. The small sample size of seven participants restricts the generalizability of the findings. Additionally, variability in participants’ familiarity with technology likely influenced their perceptions and survey responses. System performance issues, such as occasional speech recognition errors, also may have affected user satisfaction.

For future research, several improvements could be considered. Allowing users to dynamically select or adjust the robot personalities could create a more customized and engaging experience. Enhancing the system’s response latency and speech recognition accuracy would further improve interaction quality. Exploring interactions that rely solely on a language model without predefined answers could offer deeper insights into the effects of personality on engagement and learning. Testing the robots in real-world environments, such as museums, would also help assess multi-user functionality and better understand their potential to adapt to real-time data. Lastly, investigating how robots can enhance educational accessibility for underrepresented visitor groups, such as visually impaired individuals, remains an important avenue for future work.

5 Conclusion

The user test results demonstrated a generally neutral-to-positive perception of robots, with generational differences influencing preferences and engagement. Younger participants favored Furhat B for its energetic and modern style, while older participants resonated more with Furhat A’s calm personality, though they still rated Furhat B higher in usability. Furhat B excelled in response behavior and educational value, especially among younger users, confirming its broader appeal. However, the study’s small sample size and technical limitations restrict the generalization of the findings.

Future research should explore dynamic personality adjustments, real-world testing, and accessibility enhancements to improve personalization and engagement for diverse audiences.

References

[1] (2024, March) National museum and gallery visits 27 percent lower than pre-pandemic levels. [Accessed: 06-Jan-2025]. [Online]. Available: https://museumsandheritage.com/advisor/posts/national-museum-and-gallery-visits-27-lower-than-pre-pandemic-levels/

[2] G. Gaia, S. Boiano, and A. Borda, "Engaging museum visitors with ai: The case of chatbots," in *Museums and Digital Culture*, ser. Springer Series on Cultural Computing, T. Giannini and J. Bowen, Eds. Springer, Cham, 2019, pp. 309–329.

[3] W. H. Organization. (2023, Aug.) Blindness and vision impairment. [Accessed: 06-Jan-2025]. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[4] C. Coates. (2023, Dec.) Best practice in making museums more accessible to visually impaired visitors. [Accessed: 06-Jan-2025]. [Online]. Available: https://www.museumnext.com/article/making-museums-accessible-to-visually-impaired-visitors/

[5] F. Barth, H. Candello, P. Cavalin, and C. Pinhanez, "Intentions, meanings, and whys: designing content for voice-based conversational museum guides," in *Proceedings of the 2nd Conference on Conversational User Interfaces*, July 2020, pp. 1–8.

[6] A. Niculescu, B. van Dijk, A. Nijholt, H. Li, and S. L. See, "Making social robots more attractive: the effects of voice pitch, humor and empathy," *International Journal of Social Robotics*, vol. 5, pp. 171–191, 2013.

[7] M. Adcock, S. Harding, and D. Trainor, "Co-design of a voice-driven interactive smart guide for museum accessibility and management," 2023, retrieved from https://www.researchgate.net/publication/381602737_Co-design_of_a_Voice-Driven_Interactive_Smart_Guide_for_Museum_Accessibility_and_Management.

[8] D. E. Kulathunga, "Exploring the experience of different robot personalities in enhancing university students' learning," Master thesis, Tampere University, 2024, https://urn.fi/URN:NBN:fi:tuni-202405105690.

[9] J. Peponis, R. Conroy Dalton, J. Wineman, and N. Dalton, "Path, theme, and narrative in open plan exhibition settings," 2004, retrieved from https://discovery.ucl.ac.uk/id/eprint/1097/1/PTaN.pdf.pdf.

[10] Interaction flow design. [Online]. Available: https://miro.com/app/board/uXjVL6vgdak=/?share_link_id=551861596033

[11] D. S. Syrdal, K. Dautenhahn, K. L. Koay, and M. L. Walters, "The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study," in *Adaptive and emergent behaviour and complex systems*, 2009.

[12] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, "Measurement of negative attitudes toward robots," *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, vol. 7, no. 3, pp. 437–454, 2006. doi: 10.1075/is.7.3.14nom

[13] O. H. Chi, S. Jia, Y. Li, and D. Gursoy, "Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (ai) social robots in service delivery," *Computers in Human Behavior*, vol. 118, p. 106700, 2021.

[14] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, "Measurement of negative attitudes toward robots," *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, vol. 7, no. 3, pp. 437–454, 2006.

[15] M. Schrepp, H. Sandkühler, and J. Thomaschewski, "How to create short forms of ueq+ based questionnaires?" in *Mensch und Computer 2021-Workshopband*, 2021. doi: 10.18420/muc2021-mci-ws01-230

[16] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 71–81, 2009. doi: 10.1007/s12369-008-0001-3

[17] S. N. Hadie and M. S. Yusoff, "Assessing the validity of the cognitive load scale in a problem-based learning setting," *Journal of Taibah University Medical Sciences*, vol. 11, no. 3, pp. 194–202, 2016.

[18] Demonstration video of the interaction of the Furhat with a visitor. [Online]. Available: https://youtu.be/xe3Hnnjxm10

Appendix

.1 Received feedback

The feedback of the supervisor that we got on our project pre-study was very constructive. Her questions highlighted that we had to be more detailed in our writing, such as mentioning that we would be using the virtual Furhat and not the physical version, and to explain that we would vary the order of the virtual museum presentation to minimize potential bias, and made us think about some aspects that we had not think of before. Her advise on stating a concrete age to segregate the two user groups (age 30 and below, and above age of 30), was proven useful for the final analysis of the results. Overall, the feedback of our supervisor was essential to the final result and we would like to thank her for her support.

.2 Full survey for user testing

Appendix continues on the next page.

Survey for User Test

Dear Participant,

This is study about interactions with virtual museum guides, conducted as part of the DT2140 HT24 Multimodal Interaction and Interfaces course at KTH University.

Your first task is to answer questions about your demographics before you interact with the Furhat robot.

We do not collect sensitive personal data such as your name or address.
All data will be anonymized. Your participation is voluntary. You may stop participating at any time by closing the browser window or the program to withdraw from the study.
Partial data will not be analyzed.

For questions, please do not hesitate to contact any of our team members:

- Alessandro Amandonico (aleama@kth.se)
- Malin Brilon (brilon@kth.se)
- Matteo Del Prato (matteodp@kth.se)
- Stijn Teekens (teekens@kth.se)

* Gibt eine erforderliche Frage an

1. For participating in the study, please confirm that: *

   Wählen Sie alle zutreffenden Antworten aus.

   - [ ] I have read the above-mentioned conditions.

2. Please confirm that: *

   Wählen Sie alle zutreffenden Antworten aus.

   - [ ] My participation is voluntary and I know that I can abandon the survey at any point.

3. Enter the codename that was given to you by the researcher: *

Demographic Questions

4. What is your age in years? *

_____

5. What gender do you identify with? *

Markieren Sie nur ein Oval.

- ◯ Male
- ◯ Female
- ◯ Non-binary
- ◯ Prefer not to say
- ◯ Sonstiges: _____

6. What is the highest level of school you have completed? *

Markieren Sie nur ein Oval.

- ◯ Less than high school degree
- ◯ High school degree or equivalent (e.g., GED)
- ◯ Some college but no degree
- ◯ Associate degree
- ◯ Bachelor degree
- ◯ Graduate degree (e.g., Masters, PhD, M.D)

7. I know a lot about AI social service robots. *

Markieren Sie nur ein Oval.

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

8. I am familiar with AI social service robots. *

Markieren Sie nur ein Oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

9. I have much knowledge about AI social service robots. *

Markieren Sie nur ein Oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

10. I am more familiar than the average person regarding AI social service robots. *

Markieren Sie nur ein Oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

11. I would feel uneasy if I was given a job where I had to use robots. *

Markieren Sie nur ein Oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

12. If robots had emotions, I would be able to make friends with them. *

Markieren Sie nur ein Oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

13. I feel comforted being with robots that have emotions. *

Markieren Sie nur ein Oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Instruction

In this study, we are exploring the interaction with a virtual museum's guide in a virtual art gallery. You will interact with a virtual gallery, and our robot will guide you. You can ask the robot questions about the art at any time.

This research is part of the DT2140 HT24 Multimodal Interaction and Interfaces course. Your participation will help us understand how to improve user interaction and personalization in virtual museum tours.

If you do not mind, I would like to make an audio and screen recording of our session. This will allow me to review your comments later, ensuring I am not distracted by taking notes during our conversation. Your recordings will be kept confidential and used only for the purposes of this study.

1. **Interaction Flow:**
    - You will interact with two different Furhat robots, each with a unique personality and appearance.
    - First, you will interact with Furhat A, who will greet you and provide information about the virtual museum.
    - You can navigate through the virtual gallery displayed on PowerPoint and ask Furhat A questions about any art pieces you encounter.
    - Furhat A will provide answers and suggest a personalized path through the gallery based on your interests.
    - After a short break, you will interact with Furhat B following a similar procedure.
    - You will again navigate through the virtual gallery, ask questions, and receive personalized guidance based on your preferences.

Please think aloud during the study. Share your thoughts, feelings, and any difficulties or positive experiences you encounter. Your honest opinions are crucial. If you have any questions or need clarification at any point, please feel free to ask. Remember, you are free to leave the study at any time.

I am a neutral evaluator, so nothing you say will hurt my feelings. Your honest opinions and feedback are incredibly valuable and can only help improve our research. You are the expert in your own experience.

After each interaction, you will be asked to complete a brief survey to share your impressions and experiences. This will help us measure factors like satisfaction, ease of interaction, engagement, and educational value.

Ensure you are comfortable with the setup:

- You will use a computer to navigate the virtual gallery.
- You will communicate with the Furhat robots using voice commands.

I will be here to assist you if you encounter any issues or have any questions.

Thank you for your participation. Your contributions are highly appreciated and will significantly aid our research.

Has your first interaction ended?

If **yes** go further in the survey if **not** complete the first interaction.

For supervisor

14. With which Furhat has the user interacted first? *

Markieren Sie nur ein Oval.

- Furhat A (Jenna)
- Furhat B (Bill)

Questions after first interaction

Efficiency

15. To achieve my goals, I consider the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
slow	◯	◯	◯	◯	◯	◯	◯	fast

16. To achieve my goals, I consider the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
inefl	◯	◯	◯	◯	◯	◯	◯	efficient

17. To achieve my goals, I consider the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
impı	◯	◯	◯	◯	◯	◯	◯	practical

18. To achieve my goals, I consider the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
clutt	◯	◯	◯	◯	◯	◯	◯	organized

Usefulness

19. I consider the possibility of using the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
usel	◯	◯	◯	◯	◯	◯	◯	useful

20. I consider the possibility of using the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
not l	◯	◯	◯	◯	◯	◯	◯	helpful

21. I consider the possibility of using the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
not	◯	◯	◯	◯	◯	◯	◯	beneficial

22. I consider the possibility of using the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
not	◯	◯	◯	◯	◯	◯	◯	rewarding

Intuitive Use

23. In my opinion, using the product is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
diffi	◯	◯	◯	◯	◯	◯	◯	easy

24. In my opinion, using the product is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
illog	◯	◯	◯	◯	◯	◯	◯	logical

25. In my opinion, using the product is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
not	◯	◯	◯	◯	◯	◯	◯	plausible

26. In my opinion, using the product is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
inco	◯	◯	◯	◯	◯	◯	◯	conclusive

Specific to Virtual Gallery and Voice Interaction

Response behavior

27. In my opinion the response behaviour of the furhar robot is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
artif	◯	◯	◯	◯	◯	◯	◯	natural

28. In my opinion, the response behavior of the furhar robot is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
unpl	◯	◯	◯	◯	◯	◯	◯	pleasant

29. In my opinion, the response behavior of the furhar robot is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
unlik	◯	◯	◯	◯	◯	◯	◯	likeable

30. In my opinion, the response behavior of the furhar robot is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
bori	◯	◯	◯	◯	◯	◯	◯	entertaining

Response quality

31. The answers and questions asked by the furhat robot are *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
inap	◯	◯	◯	◯	◯	◯	◯	suitable

32. The answers and questions asked by the furhat robot are *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
usel	◯	◯	◯	◯	◯	◯	◯	useful

33. The answers and questions asked by the furhat robot are *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
not h	◯	◯	◯	◯	◯	◯	◯	helpful

34. The answers and questions asked by the furhat robot are *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
unin	◯	◯	◯	◯	◯	◯	◯	intelligent

Comprehensibility

35. In my opinion, the furhat robot has understood my voice commands *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
com	◯	◯	◯	◯	◯	◯	◯	simple

36. In my opinion, the furhat robot has understood my voice commands *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
amb	◯	◯	◯	◯	◯	◯	◯	unambiguous

37. In my opinion, the furhat robot has understood my voice commands *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
inac	◯	◯	◯	◯	◯	◯	◯	accurate

38. In my opinion, the furhat robot has understood my voice commands *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
enig	◯	◯	◯	◯	◯	◯	◯	explainable

Anthropomorphism

39. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Fake	◯	◯	◯	◯	◯	Natural

40. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Mas	◯	◯	◯	◯	◯	Humanlike

41. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Unc	◯	◯	◯	◯	◯	Conscious

42. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Artif	◯	◯	◯	◯	◯	Lifelike

43. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Mov	◯	◯	◯	◯	◯	Moving elegantly

Animacy

44. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Dea	◯	◯	◯	◯	◯	Alive

45. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Stag	◯	◯	◯	◯	◯	Lively

46. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Mec	◯	◯	◯	◯	◯	Organic

47. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Artif	◯	◯	◯	◯	◯	Lifelike

48. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Iner	◯	◯	◯	◯	◯	Interactive

49. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Apa	◯	◯	◯	◯	◯	Responsive

Likability

50. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Disli	◯	◯	◯	◯	◯	Like

51. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Unfr	◯	◯	◯	◯	◯	Friendly

52. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Unki	◯	◯	◯	◯	◯	Kind

53. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Unp	◯	◯	◯	◯	◯	Pleasant

54. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Awf	◯	◯	◯	◯	◯	Nice

Perceived Intelligence

55. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Inco	◯	◯	◯	◯	◯	Competent

56. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Igno	◯	◯	◯	◯	◯	Knowledgeable

57. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Irres	◯	◯	◯	◯	◯	Responsible

58. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Unir	◯	◯	◯	◯	◯	Intelligent

59. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Fool	◯	◯	◯	◯	◯	Sensible

Perceived safety

60. Please rate your emotional state on these scales *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Anxi	◯	◯	◯	◯	◯	Relaxed

61. Please rate your emotional state on these scales *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Agit	◯	◯	◯	◯	◯	Calm

62. Please rate your emotional state on these scales *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Quie	◯	◯	◯	◯	◯	Surprised

Self-Perceived Learning

63. The interaction with the Furhat robot significantly enhanced my understanding of the art pieces. *

Markieren Sie nur ein Oval.

- ◯ Not at all the case
- ◯ Somewhat the case
- ◯ Moderately the case
- ◯ Mostly the case
- ◯ Completely the case

64. The interaction with the Furhat robot significantly enhanced my understanding of the terminology associated with the art. *

Markieren Sie nur ein Oval.

- ◯ Not at all the case
- ◯ Somewhat the case
- ◯ Moderately the case
- ◯ Mostly the case
- ◯ Completely the case

65. The interaction with the Furhat robot significantly enhanced my knowledge of concepts and definitions related to the art. *

Markieren Sie nur ein Oval.

- ◯ Not at all the case
- ◯ Somewhat the case
- ◯ Moderately the case
- ◯ Mostly the case
- ◯ Completely the case

66. The interaction with the Furhat robot significantly enhanced my knowledge and understanding of the art pieces overall. *

Markieren Sie nur ein Oval.

- ◯ Not at all the case
- ◯ Somewhat the case
- ◯ Moderately the case
- ◯ Mostly the case
- ◯ Completely the case

Break

Take a 5-min break

Has your second interaction ended?

If **yes** go further in the survey if **not** complete the first interaction.

Questions after second interaction

Efficiency

67. To achieve my goals, I consider the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
slow	◯	◯	◯	◯	◯	◯	◯	fast

68. To achieve my goals, I consider the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
inefl	◯	◯	◯	◯	◯	◯	◯	efficient

69. To achieve my goals, I consider the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
impı	◯	◯	◯	◯	◯	◯	◯	practical

70. To achieve my goals, I consider the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
clutt	◯	◯	◯	◯	◯	◯	◯	organized

Usefulness

71. I consider the possibility of using the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
usel	◯	◯	◯	◯	◯	◯	◯	useful

72. I consider the possibility of using the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
not l	◯	◯	◯	◯	◯	◯	◯	helpful

73. I consider the possibility of using the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
not l	◯	◯	◯	◯	◯	◯	◯	beneficial

74. I consider the possibility of using the furhat robot as *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
not	◯	◯	◯	◯	◯	◯	◯	rewarding

Intuitive Use

75. In my opinion, using the product is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
diffi	◯	◯	◯	◯	◯	◯	◯	easy

76. In my opinion, using the product is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
illog	◯	◯	◯	◯	◯	◯	◯	logical

77. In my opinion, using the product is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
not	◯	◯	◯	◯	◯	◯	◯	plausible

78. In my opinion, using the product is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
inco	◯	◯	◯	◯	◯	◯	◯	conclusive

Specific to Virtual Gallery and Voice Interaction

Response behavior

79. In my opinion the response behaviour of the furhar robot is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
artif	◯	◯	◯	◯	◯	◯	◯	natural

80. In my opinion, the response behavior of the furhar robot is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
unpl	◯	◯	◯	◯	◯	◯	◯	pleasant

81. In my opinion, the response behavior of the furhar robot is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
unlił	◯	◯	◯	◯	◯	◯	◯	likeable

82. In my opinion, the response behavior of the furhar robot is *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
bori	◯	◯	◯	◯	◯	◯	◯	entertaining

Response quality

83. The answers and questions asked by the furhat robot are *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
inap	◯	◯	◯	◯	◯	◯	◯	suitable

84. The answers and questions asked by the furhat robot are *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
usel	◯	◯	◯	◯	◯	◯	◯	useful

85. The answers and questions asked by the furhat robot are *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
not l	◯	◯	◯	◯	◯	◯	◯	helpful

86. The answers and questions asked by the furhat robot are *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
unin	◯	◯	◯	◯	◯	◯	◯	intelligent

Comprehensibility

87. In my opinion, the furhat robot has understood my voice commands *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
com	◯	◯	◯	◯	◯	◯	◯	simple

88. In my opinion, the furhat robot has understood my voice commands *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
amb	◯	◯	◯	◯	◯	◯	◯	unambiguous

89. In my opinion, the furhat robot has understood my voice commands *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
inac	◯	◯	◯	◯	◯	◯	◯	accurate

90. In my opinion, the furhat robot has understood my voice commands *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
enig	◯	◯	◯	◯	◯	◯	◯	explainable

Anthropomorphism

91. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Fake	◯	◯	◯	◯	◯	Natural

92. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Mas	◯	◯	◯	◯	◯	Humanlike

93. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Unc	◯	◯	◯	◯	◯	Conscious

94. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Artif	◯	◯	◯	◯	◯	Lifelike

95. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Mov	◯	◯	◯	◯	◯	Moving elegantly

Animacy

96. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Dea	◯	◯	◯	◯	◯	Alive

97. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Stag	◯	◯	◯	◯	◯	Lively

98. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Mec	◯	◯	◯	◯	◯	Organic

99. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Artif	◯	◯	◯	◯	◯	Lifelike

100. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Iner	◯	◯	◯	◯	◯	Interactive

101. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Apa	◯	◯	◯	◯	◯	Responsive

Likability

102. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Disli	◯	◯	◯	◯	◯	Like

103. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Unfr	◯	◯	◯	◯	◯	Friendly

104. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Unki	◯	◯	◯	◯	◯	Kind

105. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Unp	◯	◯	◯	◯	◯	Pleasant

106. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Awf	◯	◯	◯	◯	◯	Nice

Perceived Intelligence

107. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Inco	◯	◯	◯	◯	◯	Competent

108. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Ignc	◯	◯	◯	◯	◯	Knowledgeable

109. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Irres	◯	◯	◯	◯	◯	Responsible

110. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Unir	◯	◯	◯	◯	◯	Intelligent

111. Please rate your impression of the furhat robot on these scales: *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Fool	◯	◯	◯	◯	◯	Sensible

Perceived safety

112. Please rate your emotional state on these scales *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Anxi	◯	◯	◯	◯	◯	Relaxed

113. Please rate your emotional state on these scales *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Agit	◯	◯	◯	◯	◯	Calm

114. Please rate your emotional state on these scales *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Quie	◯	◯	◯	◯	◯	Surprised

Self-Perceived Learning

115. The interaction with the Furhat robot significantly enhanced my understanding of the art pieces. *

Markieren Sie nur ein Oval.

- ◯ Not at all the case
- ◯ Somewhat the case
- ◯ Moderately the case
- ◯ Mostly the case
- ◯ Completely the case

116. The interaction with the Furhat robot significantly enhanced my understanding of the terminology associated with the art. *

Markieren Sie nur ein Oval.

- ◯ Not at all the case
- ◯ Somewhat the case
- ◯ Moderately the case
- ◯ Mostly the case
- ◯ Completely the case

117. The interaction with the Furhat robot significantly enhanced my knowledge of concepts and definitions related to the art. *

Markieren Sie nur ein Oval.

- Not at all the case
- Somewhat the case
- Moderately the case
- Mostly the case
- Completely the case

118. The interaction with the Furhat robot significantly enhanced my knowledge and understanding of the art pieces overall. *

Markieren Sie nur ein Oval.

- Not at all the case
- Somewhat the case
- Moderately the case
- Mostly the case
- Completely the case

Thank You Message

Dear Participant,

Thank you very much for your participation in our study. Your insights and feedback are incredibly valuable to us and will greatly contribute to our research on improving user interactions with virtual museum guides.

We appreciate your time and effort in helping us understand how to better design and implement such systems to enhance user experience and personalization in virtual museum tours.

If you have any further questions or would like to receive a copy of the final project results, please do not hesitate to contact us.

Once again, thank you for your participation!

Best regards,

Alessandro Amandonico (aleama@kth.se)

Malin Brilon (brilon@kth.se)

Matteo Del Prato (matteodp@kth.se)

Stijn Teekens (teekens@kth.se)

Master Students in Human Computer Interaction and Design

KTH Royal Institute of Technology

Dieser Inhalt wurde nicht von Google erstellt und wird von Google auch nicht unterstützt.

Google Formulare

.3 Table of survey layout

Before Interaction				
Topic	Question	Methode	Source	Hypothesis
Demographics	What is your age in years?	Free text	Hughes, J. L., Camden, A. A., & Yangchen, T. (2016). Rethinking and updating demographic questions: Guidance to improve descriptions of research samples. Psi Chi Journal of Psychological Research, 21(3), 138-151.	H1
	What gender do you identify with?	Male Female Non-binary Prefer not to say Other		
	What is the highest level of school you have completed?	Less than high school degree High school degree or equivalent (e.g., GED) Some college but no degree Associate degree Bachelor degree Graduate degree (e.g., Masters, PhD, M.D)		
Familiarity and experience with robots	• I know a lot about AI social service robots • I am familiar with AI social service robots • I have much knowledge about AI social service robots • I am more familiar than the average person regarding AI social service robots	5-point Likert scale, from "Strongly Disagree" to "Strongly Agree"	Chi, O. H., Jia, S., Li, Y., & Gursoy, D. (2021). Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery. Computers in Human Behavior, 118, 106700.	
NARS - Measure Attitude towards Robots	• Negative attitudes toward situations of interaction with robots ◦ "I would feel uneasy if I was given a job where I had to use robots." • Negative attitudes toward the social influence of robots ◦ "If robots had emotions, I would be able to make friends with them." (reverse scored) • Negative attitudes toward emotions in interaction with robots ◦ "I feel comforted being with robots that have emotions." (reverse scored)	5-point Likert scale, from "Strongly Disagree" to "Strongly Agree"	Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. 2006. Measurement of negative attitudes toward robots. Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems 7, 3 (2006), 437–454. https://doi.org/10.1075/is.7.3.14nom Syrdal, D. S., Dautenhahn, K., Koay, K. L., & Walters, M. L. (2009). The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. Adaptive and emergent behaviour and complex systems.	
After interaction with each furhat:				
Topic	Question	Methode	Source	Hypothesis
UEQ+	General Interaction: • Efficiency: Tasks can be finished without unnecessary effort. • Usefulness: Using the product is beneficial. • Intuitive Use: The product can be used immediately without any training or help. Specific to Virtual Gallery and Voice Interaction: • Response Behavior: The voice assistant behaves respectfully and is trustworthy. • Response Quality: The responses cover the user's information needs. • Comprehensibility: The voice assistant understands the users' instructions and questions using natural language.	Uses of semantic differential method with opposite pairs of adjectives (e.g., human-like vs. machine-like) in a seven sub-scale	Schrepp, M., Sandkühler, H., & Thomaschewski, J. (2021). How to create short forms of UEQ+ based questionnaires?. Mensch und Computer 2021-Workshopband. DOI:10.18420/muc2021-mci-ws01-230. Show in the MCI Digital Library	H2: Response Behavior, Response Quality, Comprehensibility H3: Efficiency, Usefulness H4: UEQ+ General Usefulness
GODSPEED	• Anthropomorphism (human-likeness) • Animacy (being sentient or alive) • Likeability • Perceived intelligence • Perceived safety	Uses a semantic differential method with opposite pairs of adjectives (e.g., human-like vs. machine-like) in a five sub-scale	Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics 1, 1 (2009), 71–81. https://doi.org/10.1007/s12369-008-0001-3	H1: Likability H2: Anthropomorphism, Animacy H5: Likeability, Perceived intelligence, Perceived safety
Self-Perceived Learning	• The interaction with the Furhat robot significantly enhanced my understanding of the art pieces. • The interaction with the Furhat robot significantly enhanced my understanding of the terminology associated with the art. • The interaction with the Furhat robot significantly enhanced my knowledge of concepts and definitions related to the art. • The interaction with the Furhat robot significantly enhanced my knowledge and understanding of the art pieces overall.	[ ] Not at all the case [ ] Somewhat the case [ ] Moderately the case [ ] Mostly the case [ ] Completely the case	Hadie, S. N., & Yusoff, M. S. (2016). Assessing the validity of the cognitive load scale in a problem-based learning setting. Journal of Taibah University Medical Sciences, 11(3), 194-202.	H5

Figure 9: Table of survey layout

.4 Consent form

Appendix continues on the next page.

Adaptive AI Museum Guides: Enhancing User Interaction with Furhat Robots

Consent form DT2140 HT24 Multimodal Interaction and Interfaces

Research Institution

KTH Royal Institute of Technology

Project Supervisor

Charlotte Stinkeste

csti@kth.se

Purpose of the Study

We are conducting an academic study as part of our coursework at KTH Royal Institute of Technology. The aim of this study is to gain insights into the interaction with a virtual museum guide in a virtual art gallery, focusing on the effectiveness of the robot in delivering personalized and engaging educational content. This user study will explore your experiences and perceptions during the interaction with the Furhat robots in the virtual gallery

---

Procedure

This user test will take approximately 30 to 40 minutes. We will start with the interaction of the vitual gallery and the Furhat robots, followed by a short survey. If anything is unclear, please let us know, and we will provide further clarification.

Voluntary Participation

Your participation in this user study is entirely voluntary. You have the right to withdraw from the user study at any point, without any consequence or explanation. You may also refuse to answer any specific question during the user study.

Risk and Benefits

No specific risks or benefits from participating in this user study have been identified.

Confidentiality

The information shared during the user study will be used solely for academic purposes and will remain confidential unless you provide explicit permission to share specific information. The data will be anonymized via a participant number. The data will be securely stored on KTH servers and will not be shared with third parties without your explicit consent.

I agree to the interaction being audio-recorded for transcription purposes.
☐ Yes
☐ No

I agree to the interaction being screen-recorded for research purposes.
☐ Yes
☐ No

Use of Data

The data gathered from this user study will be analyzed as part of our Multimodal Interaction and Interfaces project. It will not be shared with third parties, and only the research team will have access to the full data.

Consent

By signing below, you acknowledge that you understand the purpose of this user study and agree to participate voluntarily. You confirm that you are aware of your rights to withdraw from the user study at any time and that your confidentiality will be maintained unless otherwise agreed.

***Your Signature**:*

***Date**:*

***Researcher's Signatures**:*

Contact Information of the Team

If you have any questions or concerns regarding this user study or the study, please feel free to contact us:

- Alessandro Amandonico (aleama@kth.se)
- Malin Brilon (brilon@kth.se)
- Matteo Del Prato (matteodp@kth.se)
- Stijn Teekens (teekens@kth.se)

.5 Results and Analysis

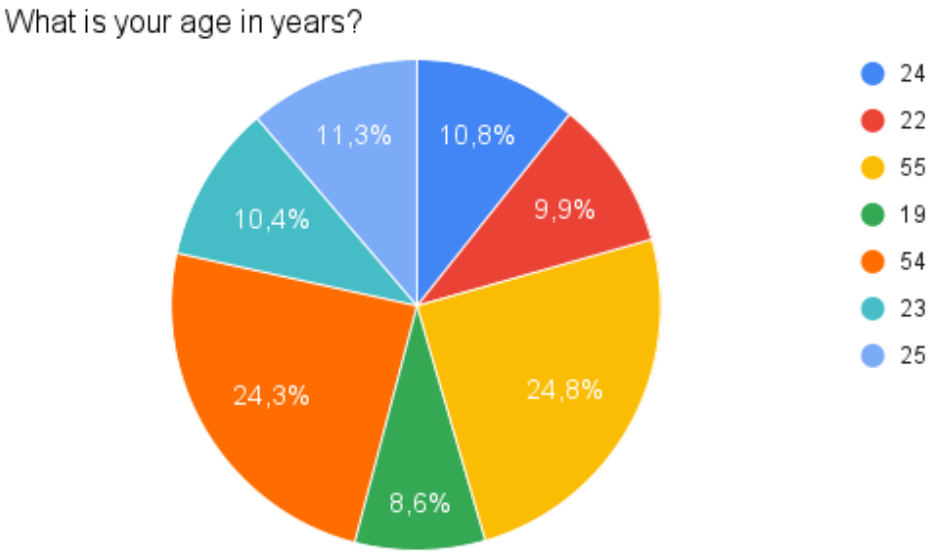

Figure 10: Age of participants

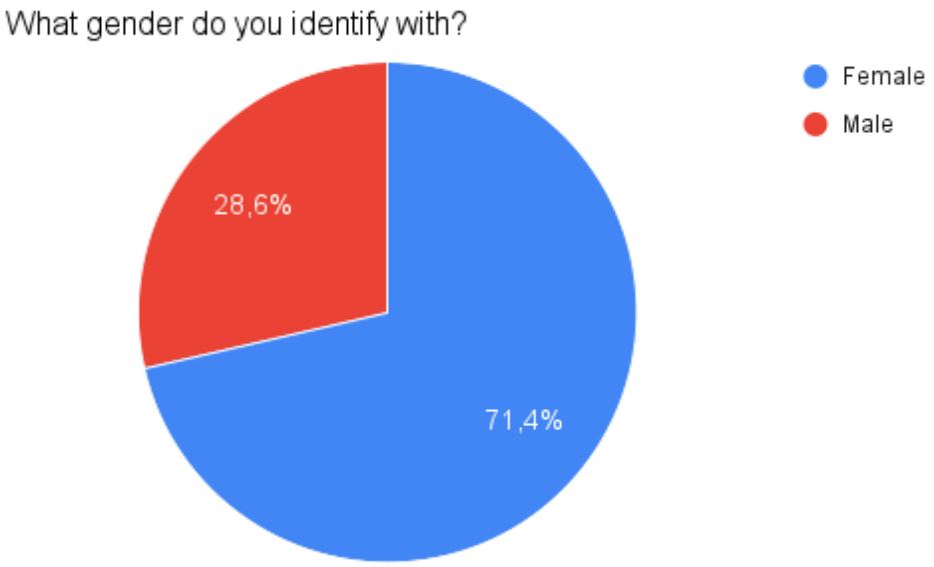

Figure 11: Gender of participants

Question Number	Question Type
Q1	I know a lot about AI social service robots.
Q2	I am familiar with AI social service robots.
Q3	I have much knowledge about AI social service robots.
Q4	I am more familiar than the average person regarding AI social service robots.

Table 2: List of ”Familiarity and experience with Robots” questions

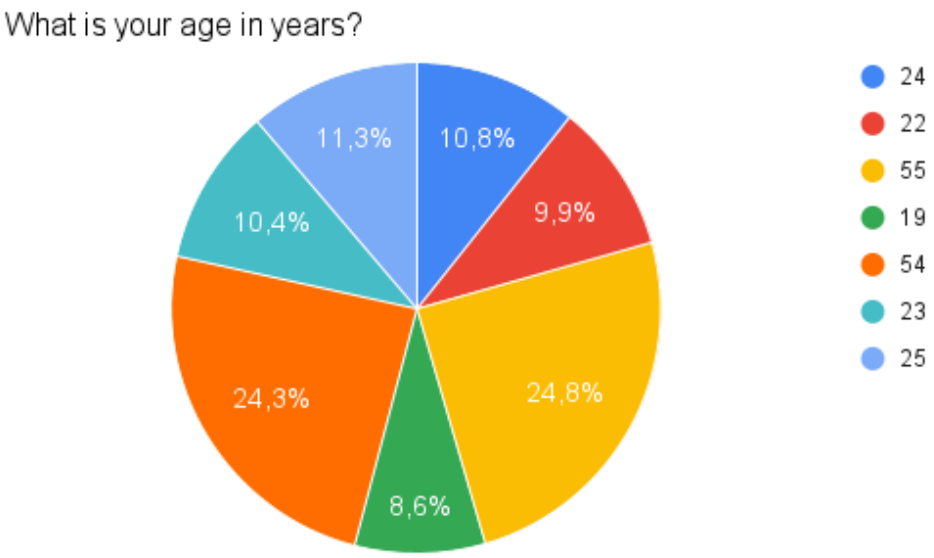

Figure 12: Educational background of participants

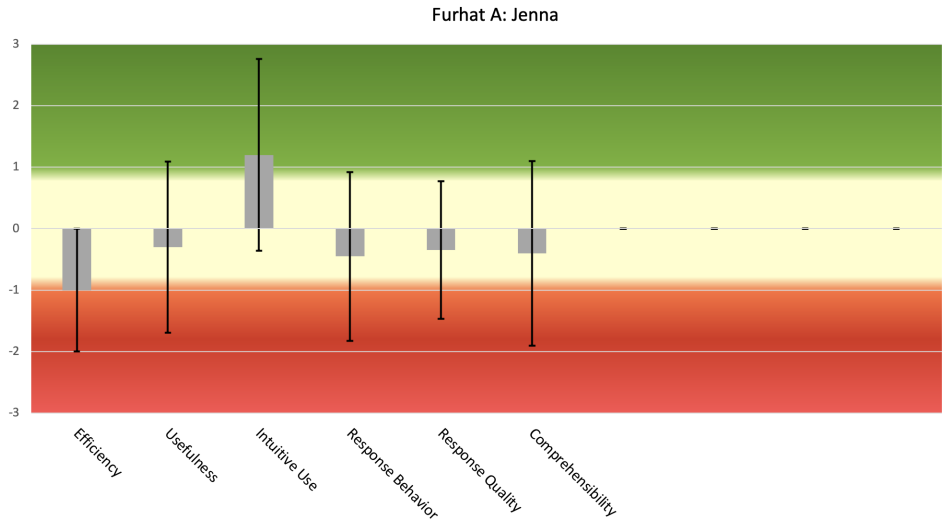

Figure 13: Furhat A UEQ+ overall results from Cluster 1

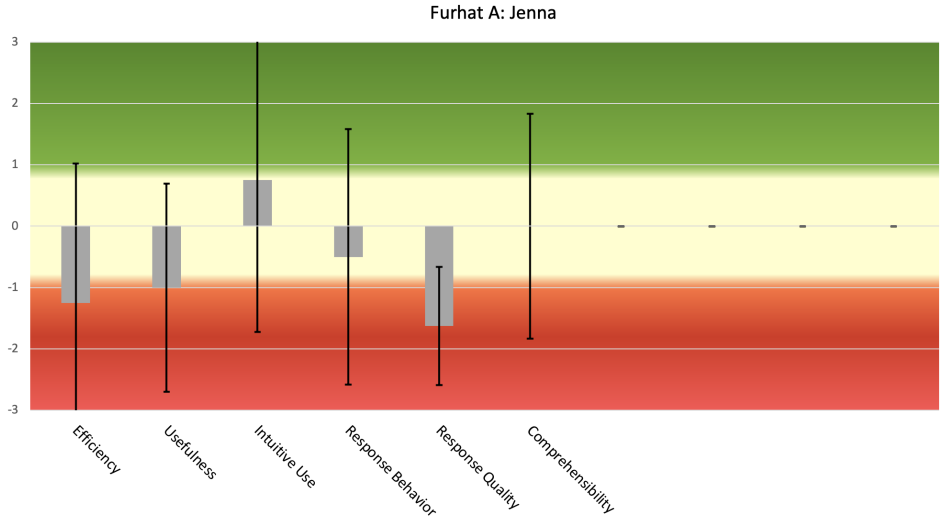

Figure 14: Furhat A UEQ+ overall results from Cluster 2

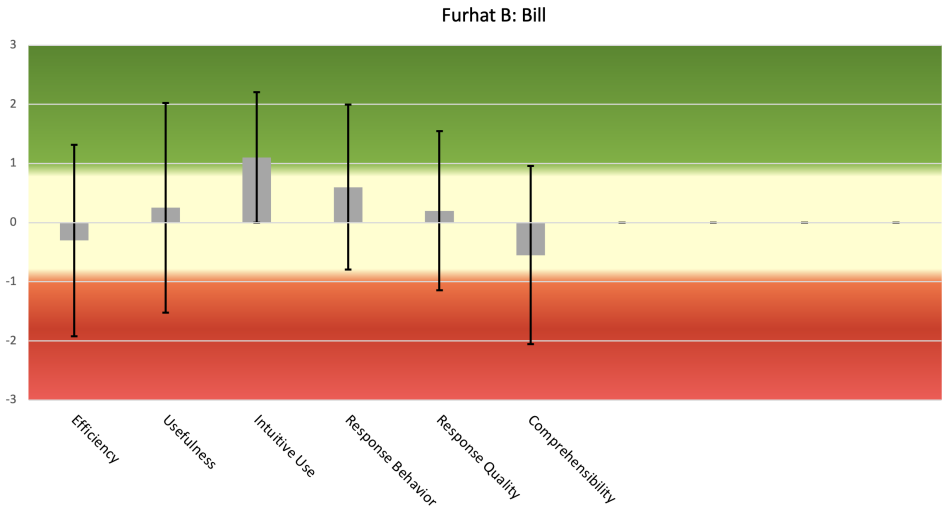

Figure 15: Furhat B UEQ+ overall results from Cluster 1

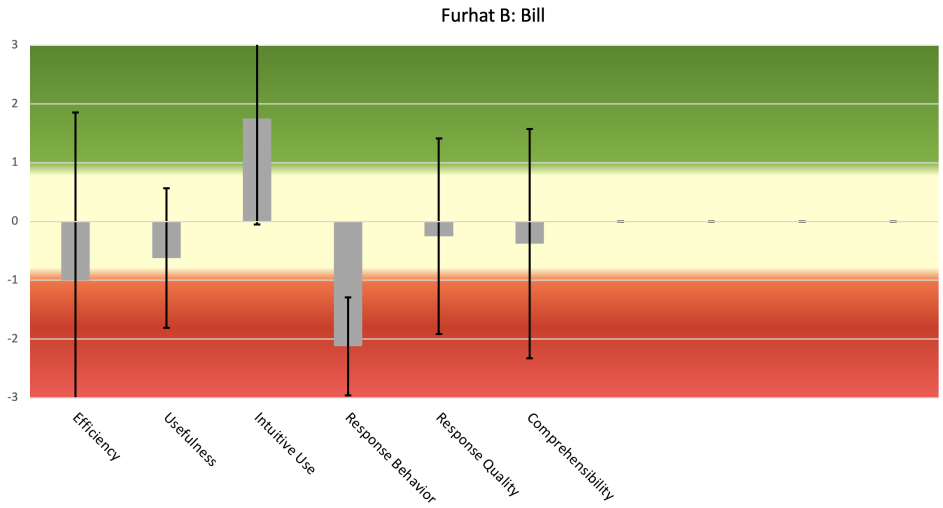

Figure 16: Furhat B UEQ+ overall results from Cluster 2

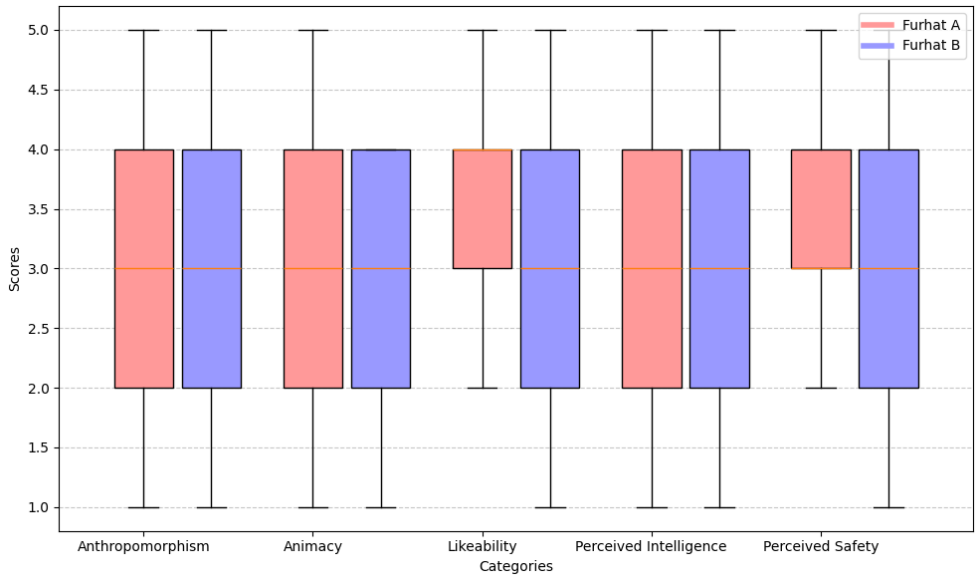

Figure 17: Furhat A vs Furhat B Overall Comparison based on GODSPEED results

Question Number	Question Type	Sub-Scale
Q1	I would feel uneasy if I was given a job where I had to use robots.	Sub-Scale 1
Q2	If robots had emotions, I would be able to make friends with them.	Sub-Scale 2
Q3	I feel comforted being with robots that have emotions.	Sub-Scale 2

Table 3: List of ”NARS - Measure Attitude towards Robots” questions

Type of Answer	Assigned Value
Strongly Disagree	1
Disagree	2
Neutral	3
Agree	4
Strongly Agree	5

Table 4: Questionnaires Likert Scale