Exposing the Brain of a Robot

DAVID GALATI, University of Twente, The Netherlands

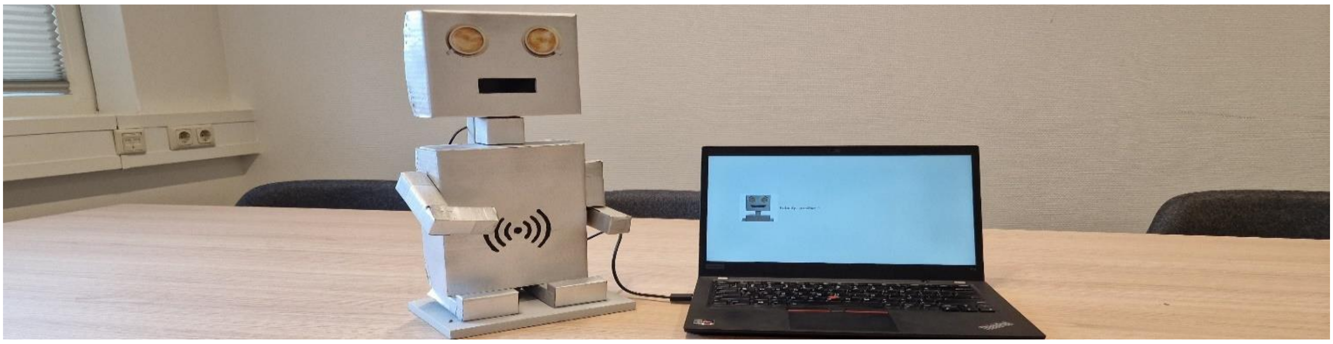

Figure 1 CoffeeBot setup during the of the experiment

Abstract - Transparency in robot and artificial intelligence cognitive processes is critical for developing trust and promoting effective human-robot interactions. By studying the link between visualization methods and the perceived traits of robots, this research aims to improve our understanding of how people's perception of robots is connected to visual cues. Participants in the experiment were asked to interact with various visualization interfaces, such as graphical user interfaces (GUI) and plain-text visualizations. The results indicate that GUIs led to higher scores in perceived intelligence, likability, and understandability compared to plain-text visualizations. Statistical analysis revealed a significant difference in understandability scores between the two visualization types, however, while perceived intelligence and likability showed positive trends with GUIs, the differences were not statistically significant. These findings emphasize the potential of GUIs in enhancing users' understanding of the robot's behavior and decision-making processes. This suggests the need for further exploration to uncover additional factors influencing these aspects and to optimize graphical visualization techniques.

Additional Key Words and Phrases: transparency, human-robot interaction, visualization, XAI, anthropomorphism

1 INTRODUCTION

Studies have revealed that many people still do not fully comprehend the potential of robots despite the fact that these technologies are being used in a variety of applications[5]. This misalignment of knowledge can cause misunderstandings, unrealistic expectations, and even distrust towards robots and AI systems. To promote safe and effective human-robot interactions, it is important to acknowledge this lack of understanding and providing individuals with insights from the functioning and decision-making processes of these technologies. Using transparency, we can align people's expectations to the reality of robot capabilities, mitigating potential errors, accidents, and harm that may arise from misinterpretations of robot behavior.

1.1 Defining transparency

Defining transparency has been a challenge in related literature. According to [8], transparency can be seen in different ways depending on the topic , with different definitions from both a legal standpoint and the point of view of different stakeholders. Transparency in the behavior of robots and AI is essential for building trust, improving safety, and enabling effective human-robot interactions. Theodorou et al [17] also identifies multiple dimensions of transparency, such as the *lack of deception* or *mechanisms for reporting reliability* and *decision making*. In this research, we will use Theodorou et al's [17] definition of transparency as *a mechanism that exposes decision making*, to aid people's understanding of the robot. As a result, better ways for displaying the behavior of robots and AI are required to make these technologies clearer and more understandable to people.

1.2 Problem Statement

As robots and AI systems become more and more present in various domains, the lack of openness in their behavior and decision-making processes is becoming increasingly problematic.

The complexity of robotic systems, especially those based on machine learning algorithms, has brought up the "black box" problem. Even the creators of these systems may struggle to provide comprehensive explanations due to their probabilistic nature. The lack of transparency in machine learning systems has raised ethical concerns [15]. The need for explainable artificial intelligence (XAI) and various strategies to increase transparency in machine learning systems have been addressed in a number of academic papers, including Arrieta et al. [1]. This ambiguity not only causes confusion but also hinders understanding robot behavior. Since the robot is focused on dialogue, adding an audio transparency interface on top of the existing dialogue might prove overwhelming and hard to follow [20], so this study will focus on visual methods of transmitting information.

This, in turn, will help to improve human-robot interaction, which is a vital part of robotics research. Therefore, more research is needed to identify and evaluate visualization methods that are effective in transmitting information about a robot's behavior. This will be helpful in ensuring that the communication from robots to humans is clear and concise, hence promoting the creation of reliable and trustworthy robots.

1.3 Goals

The goal of this study was to check the effectiveness of various methods of visualizing the behavior of robots and AI and to examine how different ways of this information are associated with a better understanding of robot behavior by humans. This research was based on an experiment in which participants interacted with a talking robot that exhibited different levels of transparency. This was done to evaluate participants' initial expectations of a talking robot and contrast them with their impressions following interactions with the robot using various visualization interfaces. The interfaces offered insights into the data the robot had access to and how it was interpreted. The study also aimed to investigate the effects of providing participants with insights into the robot's cognitive processes on their perceptions of the robot's intelligence and likeability. Limited information about the robot's pseudo-emotions was presented for this purpose using a graphical user interface. By examining the impact of different visualization techniques on participants' understanding of robot behavior and their assessments of the robot's intelligence and likability, this research contributed to a deeper understanding of human-robot interactions. The findings highlight the potential benefits and implications of transparency and visualization in enhancing the interpretability of robot behavior and fostering effective human-robot communication.

2 RELATED WORK

The works of Theodorou, Felzman, Wortham et al[8,16,17,21], have not only tackled the importance of real-time visualization interfaces, but also provided an insight into the process of developing such interfaces. This is linked to many other HRI subjects that helped shape this study. In this section I will group the literature review in three relevant levels.

2.1 Transparency

Although initially the aim of the experiment was to focus on visualizing robot decision-making, the very nature of our robot, a talking robot that specializes in question generation made it difficult to find any relevant behavioral elements to visualize. Dietrich et al, [6] however showcased a robot arm visualization that focuses on perception and awareness rather than behavior, suggesting that not only behavioral information, but also perceptual data is important for improving the transparency of robots This paper will be focusing on the effects of visualizing robot's perception.

2.2 Expectations

Understanding what humans expect of social robots before interacting with them is relevant to our study to set a baseline for how humans perceive different traits of a robot.

Minae Kwon did extensive research in the field of HRI, especially concerning people's expectations of different social robots. People sometimes have much higher expectations of social robots, [10,12], at the same time if a robot cannot do a certain thing, how should that information be conveyed to the people, and what is the next best thing it can do [11]?

On top of this, people's expectations of robots depends on their culture, Europeans tend to have less exposure to robots, and different expectations than the Japanese, for example [4,9]. Those studies suggested that previous expectations of robots are important in interpreting the results this experiment.

2.3 Anthropomorphism and Likability

Anthropomorphism is the interpretation of nonhuman things or events in terms of human characteristics [22]. It was showed that anthropomorphism affects empathy towards robots [13], and that anthropomorphism is linked to an increase in perceived intelligence [7]. Displaying signs of emotion affects the perceived anthropomorphisms [14]. This has led the question of whether or not associating human-like traits such as emotional reactions to words to a robot increases the likability and perceived intelligence of the robot. This theory is the reason for including traits such as intelligence and likability in the dependent variables of this research.

3 RESEARCH QUESTIONS AND HYPOTHESES

3.1 Research questions

RQ 1: How do people's expectations of a talking robot's intelligence based on description and looks compare to their impressions after the interaction?

RQ 2: How does visualizing speech data available to a social robot influence its perceived traits, with a blind interaction[1] as baseline:

- *RQ 2.1: How does a plain-text visualization of a robot's perception compare* to *a graphical visualization in terms of understandability?*
- *RQ 2.2: How does a plain-text visualization of a robot's perception compare to a graphical visualization in terms of perceived intelligence?*
- *RQ 2.3: How does a plain-text visualization of a robot's perception compare to a graphical visualization in terms of likability?*

3.2 Hypotheses

Based on the insights gain from literature from Section 2, this study also aims to examine the impact of emotions and anthropomorphism on the perceived intelligence of the robot. We hypothesize the following:

> H1: *People's expectations of the robot's intelligence will be higher than its perceived intelligence scores after the interaction.*
>
> H2: *Any kind of visualization interface will lead to an increase in understandability of the robot.*
>
> H3: *After being exposed to a graphical user interface that associates emotion icons with the emotions extracted from sentiment analysis, participants' ratings of both the likeability and the perceived intelligence of the robot will increase.*

H3 is based on the assumption that the visual representation of emotions will evoke a greater sense of anthropomorphism.

Through our experiment and analysis, we have investigated these hypotheses to gain a better understanding of how emotions and anthropomorphism are linked to the perceived traits of robots.

[1] interaction without visualizing any interface.

4 METHODOLOGY

The methodology used for this research is structured in three main sub-sections: participants, materials, and experiment design. More information about each section will be described in detail bellow.

4.1 Participants

The participants of this study were all university students. A total of 18 students (12 male, 6 female) from technical studies, including computer science and various engineering disciplines, have been recruited by word of mouth and have participated in this study. Participants were between the ages of 20 and 24, with 21 being the median age. No compensation of any kind was offered in exchange. Recruiting university students as participants offers advantages such as accessibility, in such a short time span. Before participating, participants were provided with information about the study's objectives and methods and were given an opportunity to ask any questions they had before giving their informed consent.

4.2 Materials and setup

Given the time constraints of this study, we utilized a simple robot named CoffeeBot [19], which was specifically developed for investigating conversational question generation. CoffeeBot ran a cloud-based dialogue engine, and Flipper 2.0 [18] as the underlying technology, that would build questions whose aim were to better get to know someone, questions' topics ranged from weather, personal preferences to pop-culture and preferred artists or TV shows. The dialogue engine would deliver the questions, while the software ran locally by the robot would use Google Cloud API calls to employ ASR (active speech recognition) and TTS (text to speech), although the questions were mostly based on templates, the answers of the participants would be run through syntactical analysis, such that the robot could ask follow-up questions based on the answers of the participants, i.e. the robot would identify the subject of a sentence and base its next question on that subject. The robot was made from a cardboard body, which would encapsulate a Bluetooth speaker and an RFID sensor. The local software ran on a laptop that was connected to the robot via USB and Bluetooth.
The questions asked by the robot were varied for all the participants but were generated based on a list of a 100 starter questions. After some of the questions, the robot would generate follow-up questions, such as:

Q: What did you do on your last vacation?
A: I went to Barcelona
Q: What do you think about Barcelona?

In this example, the robot used syntax analysis to identify that *Barcelona* was a noun and a modifier, so it chose it as the topic of the next question, such an interaction felt natural to the participants, however, in order to add a layer of consistency, in the robot, the robot's code was tweaked in such a way that it would not care if a question had already been asked and it would not have a minimum time passed before a follow-up question would be asked. This caused a bug that caused the robot to ask the same question multiple times such as:

Q: What's your dream job?
A: I don't know
Q: Okay. We chattered about knowing this afternoon. Do you knowing?
Q: We spoke about knowing later. When do you knowing?
Q: We chattered about knowing this afternoon. How do you knowing?

Since the participants had access to the syntactic and part-of-speech analysis both in the TUI and the GUI (Figures 2 and 3), this was a good way to identify how well people understand the robot when it is misbehaving. In this case, it was expected that people would identify the issue as something related to syntactic analysis or the sentence-building algorithm. We made sure that for all participants, at least some sort of repetition or poor word parsing on take place, such that the experiment would stay consistent.

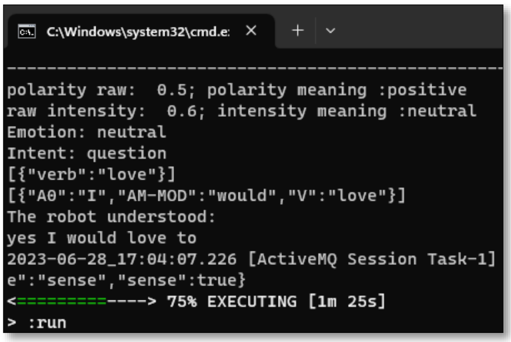

Figure 2 Instance of the TUI

Figure 3 Instance of the GUI

Figure 4 Icons reflecting the polarity of the sentiment analysis of ASR text.

4.3 Experiment Design

This study used a mixed factorial experimental design to investigate the effects of one independent variable – the visualization interface on three other dependent variables:

- Understandability
- Likability
- Perceived intelligence

There are multiple reasons for employing such an experiment design. Firstly, by using the same baseline for both visualization interfaces, the mixed factorial design allowed me to control for individual differences among participants. This helps ensure that any observed differences between the two interfaces can be attributed to the manipulation itself rather than individual variations. Also, due to the very short time span of the study, we could not expect for a large number of participants, so the mixed factorial design seemed like the best option since we can extract more information from the same number of participants. Mixed factorial design experiments also have increased statistical power, when compared to other designs such as the between-subjects design, and a within-subjects design would be prone to learning effects.

Table 1

Within-Subjects Factors		
Measure	Time	Dependent Variable
Intelligence	1	noViz_INTAvg
	2	viz_INTAvg
Likability	1	noViz_LKBAvg
	2	viz_LKBAvg
Understandability	1	noViz_understandibility
	2	viz_understandibility

Between-Subjects Factors		
	Value Label	N
Interface	TUI	9
	GUI	9

The experiment included a within-subjects factor, Time, which in this case is represented as no visualization at *Time 1*, and an interface at *Time 2*. The between-subjects factor was represented by the type of interface participants have interacted with at *Time 2.* Likability and perceived intelligence have been also assessed at *Time 0,* so before any kind of interaction, to assess people's expectations of the robot. The expectations were measured to answer RQ 1 and were analyzed in relation to *Time 1* and *Time 2* in the statistical analysis.

4.4 Experiment Procedure

After signing the informed consent form, the participants have been briefed with information about CoffeeBot, what it is, what it can do, and what it was used for. Based on this description and based on the looks of the robot, the participants were asked to fill in a questionnaire comprised of the likability and perceived intelligence. This form's purpose was to evaluate the participants' expectations.

After a dialogue with the robot, participants were asked to fill in the same form again and were asked questions to assess the understandability of the robot. Because of the limited capabilities of the robot, it was expected that people would evaluate the perceived intelligence as lower. After this interaction, people were asked to again engage in dialogue, but this time, half of the people have been presented with a terminal-based text user interface showcasing information about the text-to-speech, active speech recognition and other sensorial data available to the robot, and the other half of the people have been presented with a GUI that contained the same data, but this time, an icon that would represent the emotion of the robot extracted from the sentiment analysis of the text (Figures 3 and 4). Once again, people were asked questions to assess their understandability, but this time, after the interaction with the UI their group was assigned to.

4.5 Measures

To answer the research questions, we needed to measure the dependent variables of the experiment. To assess the perceived intelligence, likability and their prior expectations, the corresponding sections of the Godspeed Questionnaire have been used[2,3].

A semi-structured interview approach was used to evaluate the robot's understandability. After each interaction, participants were queried about their perceptions of various aspects of the robot's operation. They were specifically asked about their understanding of the robot's behavior in situations where the robot asked follow-up questions or encountered malfunctioning states, as described in Section 4.2. Participants were encouraged to provide their opinions on the reasons behind such behavior. A coding methodology was used to evaluate the interview responses. Because each dialogue was unique, participants' responses were assessed based on their alignment with the robot's underlying mechanisms. If participants provided answers that included the keywords used by the robot for syntax and part-of-speech analysis, a score of 2 out of 2 indicated a thorough understanding. In cases where participants provided answers related to issues that could have caused the problem but were not the cause in the specific context, a score of 1 out of 2 points was assigned. If the participants answered that they didn't know, or a completely incorrect answer, 0 out of 2 points were assigned. A maximum of 4 points would be awarded for 2 fully correct answers. This coding methodology facilitated the evaluation of participants' understandability of the robot while accounting for the variations in the dialogues conducted with each participant.

4.6 Analysis

The data collected from the participants was analyzed using a combination of descriptive and inferential statistical techniques. Descriptive statistics were used to summarize and present data such participants expectations of intelligence and likability. Mean scores and standard deviations were assigned to the dependent variables, which included perceived intellect and likability, of the robot's actions.

Inferential statistics were used to investigate the effects of the visualization interfaces on the dependent variables. A repeated measures ANOVA was used to compare perceived intelligence, likability, and understandability scores between the two visualization conditions, those being the terminal-based text user interface and graphical user interface.

The SPSS software was used for the statistical analysis, and the significance level was set at p 0.05 to determine statistical significance. The findings of the analyses provide insights into the impact of the visualization interfaces on participants' perceptions and understanding of the robot's behavior.

5 RESULTS

Based on the qualitative results from the Godspeed Questionnaire[3], and the understandability assessment interview, we can now answer the research questions:

- *RQ 1: How do people's expectations of a talking robot's intelligence based on description and looks compare to their impressions after the interaction?*

Before any interaction with the robot, based only on its looks and on description from Appendix B, the participants evaluated the robot's intelligence on average as 3.21 points (SD = 0.73). After the first interaction, the participants have scored the robot's intelligence with a mean of 2.81 (SD = 0.82). Both in relation to the blind interaction, and the visualization interfaces, the expectation of intelligence had p-values of less than 0.05, showing that the test is statistically significant.

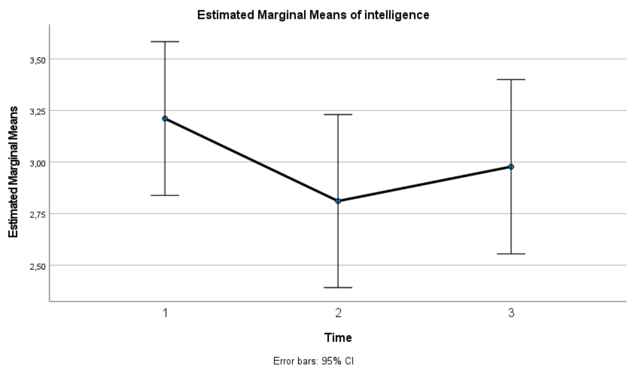

Figure 5

- *RQ 2.1: How does a plain-text visualization of a robot's perception compare to a graphical visualization in terms of understandability?*

An increase in in the understandability scores was observed with both participant groups. From a combined mean score of 1.50 (SD = 1.04) for no visualization, the understandability increases to a mean of 2.55 (SD = 0.84) for the TUI and 3.33 (SD = 0.70) for the GUI. The statistical significance (p-value) was computed as less than 0.001 for both the between-subjects and the within-subjects test, showing that seeing data about speech makes most people rate the robot as an actor that has some level of understanding.

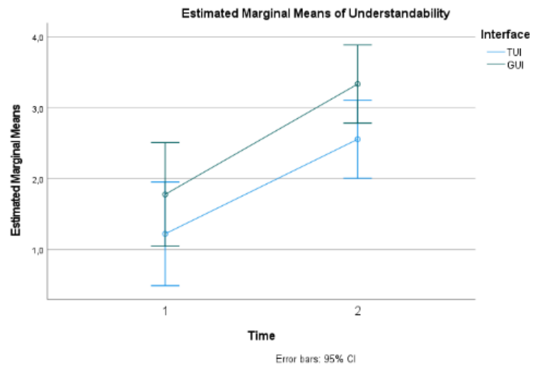

Figure 6

- *RQ 2.2: How does a plain-text visualization of a robot's perception compare to a graphical visualization in terms of perceived intelligence?*

A minor increase in the perceived intelligence has been recorded. From a combined mean score of 2.8 (SD = 0.81) for no visualization, the mean perceived intelligence has been measured at 2.91 (SD = 0.84) for the TUI and 3.04 (SD = 0.85) for the GUI. Although the values are indeed higher, the statistical significance of the ANOVA test showed p-values of 0.759 for the between-participants test, which deem the difference not statistically significant.

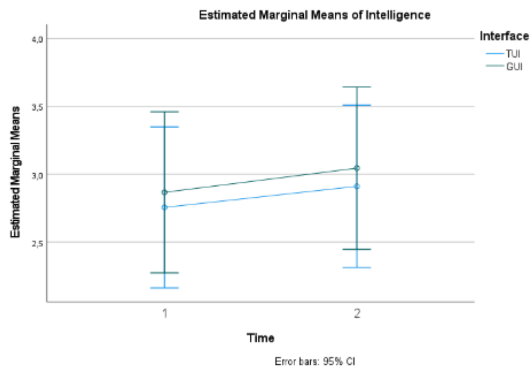

Figure 7

- *RQ 2.3: How does a plain-text visualization of a robot's perception compare to a graphical visualization in terms of likability?*

With a combined no-visualization mean score of 3.61 out of 5 points on the likability Likert scale, the mean score given by the participants after they saw the TUI changed to 3.56 (SD = 0.76), while the participants that saw the GUI scored the robot's likability at an average of 3.64 (SD = 0.76). The statistical test showed a p-value of 0.640 the between-participants test, suggesting that this has no statistical significance.

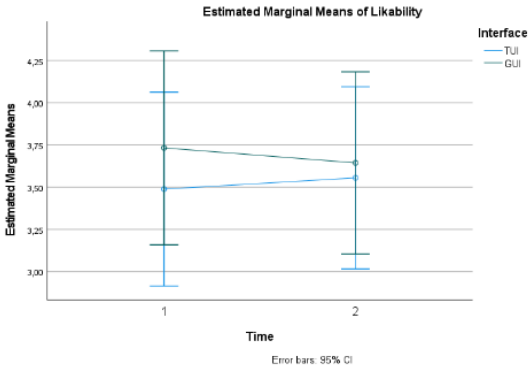

Figure 8

Although all variables seem to show an increase with respect to the time of the measurement, only the understandability and the difference between expectations and impressions scores seem to show statistical significance.

As for the understandability of the robot, besides the quantitative analysis, a quantitative analysis was conducted. A significant improvement has been observed in the answers. Specifically, they were able to recognize errors stemming from Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) processes. Most of the interview participants have also identified that the robot's follow-up questions were generated based on previous responses and predefined templates. In the absence of interfaces, most common wrong assumptions were that the robot uses some language model, the robot has a hardcoded question set and it doesn't listen to the user at all. When asked how the robot picks which word of the sentence it should use in the follow-up question, most of the participants said that the first word is used, although in both participant groups, the interfaces they were presented with showed the syntactical analysis, suggesting the fact that the word is picked based on the part-of-speech analysis.

These findings indicate that participants might have gained understanding of the inherent mechanisms and processes involved in the robot's behavior. The ability to recognize and articulate these aspects suggests a level of comprehension and engagement with the experimental setup.

6 DISCUSSION

This study's aim was to investigate the effectiveness of different methods of visualizing robot behavior and their impact on humans' understanding of robot's capabilities. The results from the previous section provide practical insights into the research questions and allow for a comprehensive discussion related to the hypotheses.

- *H1: People's expectations of the robot's intelligence will be higher than its perceived intelligence scores after the interaction.*

This hypothesis has been met since there has been a significant decrease in the perceived intelligence scores after the blind interaction. This suggests that at least in the case of CoffeeBot, people had higher expectations than their impressions after the interaction.

- *H2: Any kind of visualization interface will lead to an increase in understandability of the robot.*

As seen in (Figure 6) and backed by the results of the analysis of variance, there is a significant difference in the means of the measured understandability levels between the blind interactions, and the understandability scores of the people that interacted with the TUI and the ones that interacted with the GUI, the qualitative analysis hints the same thing.

- *H3: After being exposed to a graphical user interface (GUI) that associates emotion icons with the emotions extracted from sentiment analysis, participants' ratings of both the likeability and the perceived intelligence of the robot will increase.*

Although the statistical tests hint an difference for the between-participants means of the perceived intelligence and likability scores on the Godspeed Questionnaire[3], the which would reject this hypothesis.

6.1 Limitations

Despite the study's thoroughness and comprehensiveness, certain limitations should be addressed. To begin, the interviews done following each interaction were not recorded due to data protection concerns and scheduling restrictions. As a result, an inter-rater reliability study of the interview data was not possible. Despite efforts to maintain uniformity in the coding process, the lack of multiple raters creates a potential constraint in the dependability of the qualitative assessments.

Secondly, the study sample consisted of a relatively small number of participants (N=18), the majority of whom were technical students between the ages of 20 and 24. While this participant group allowed for in-depth exploration of the research questions within a specific demographic, it may limit the generalizability to a larger population. The participants' common background and age range may introduce a potential bias in their perspectives and responses, emphasizing the need for caution when extrapolating the results to other populations with different characteristics.

Furthermore, the experiment's ecological validity could be viewed as a limiting factor. The laboratory setting and the specific context of the robot interactions may differ from real-world scenarios in which people encounter robots and AI technologies in their daily lives. Participants' perceptions and behaviors may be influenced by the controlled environment and predefined tasks, potentially deviating from natural human-robot interactions, so, when generalizing the findings to real-world contexts, care should be taken.

These limitations should be taken into account when interpreting the results of this study, but despite these limitations, this research contributes to the field of human-robot interaction by highlighting the importance of transparency in improving the understanding between humans and robots. The findings may have significance for the design and development of robots and AI systems that communicate their activities to people effectively. We can prepare the path for the safe and effective integration of these technologies into various spheres of society by addressing the limitations of transparency and working toward more extensive and intelligible communication between people and robots.

Future research can continue by expanding the participant pool to include a more diverse range of demographics and exploring the ecological validity of human-robot interactions in real-world contexts. Additionally, the validity and reliability of qualitative evaluations can be further improved by technological and data gathering developments. We can unleash the full potential of robots and AI to help people and society as a whole if we keep working to improve transparency, understanding, and trust in human-robot interactions.

7 CONCLUSION

This research paper has investigated the effectiveness of visualization interfaces for improving the transparency and understanding of a robot's perception in a dialogue-based interaction. This study addressed the issues related to the understanding of robots' capabilities and limitations, which can lead to misunderstandings and unrealistic expectations. By exploring levels of transparency and its significance in human-robot interaction, this study sought to identify ways which factors are linked to a better perception of robots, so that people can have more pleasant experiences when interacting with machines. The methodology used in this study involved assessing how participants perceive a robot's likability, intelligence and how well they understand it. By testing the effects of two different visualization interfaces of the robot's perception, a GUI and textual interface, this study established that people had higher expectations than their impressions after the interaction, and that when presented with a GUI people are more likely to understand a robot, compared to only visualizing a TUI. A link between perceived intelligence, likability, and the level of perceived anthropomorphism of the robot was also studied, however, no statistically significant results have been found. Further research. While this study has some limitations, it represents a small step in better understanding the fast-growing world of robotics and AI, that needs to be carefully studied as robots are becoming more prevalent than ever in our daily lives.

ACKNOWLEDGMENTS

I would like to thank my research supervisors, Professor Bob R. Schadenberg and Professor Mariët Theune for their support throughout this project and Jelte van Waterschoot for his support with the CoffeeBot software.

REFERENCES

[1] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, Francisco Herrera, Francisco Herrera, and Francisco Sotelo Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion* 58, (June 2020), 82–115. DOI:https://doi.org/10.1016/j.inffus.2019.12.012

[2] Christoph Bartneck. 2023. Godspeed Questionnaire Series: Translations and Usage. . 1–35. DOI:https://doi.org/10.1007/978-3-030-89738-3_24-1

[3] Christoph Bartneck, Dana Kulic, Elizabeth A. 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 (January 2009), 71–81. DOI:https://doi.org/10.1007/s12369-008-0001-3

[4] Christoph Bartneck, Tomohiro Suzuki, Takayuki Kanda, and Tatsuya Nomura. 2006. The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI & Soc* 21, 1–2 (November 2006), 217–230. DOI:https://doi.org/10.1007/s00146-006-0052-7

[5] Elizabeth Cha, Anca D. Dragan, and Siddhartha S. Srinivasa. 2015. Perceived robot capability. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 541–548. DOI:https://doi.org/10.1109/ROMAN.2015.7333656

[6] André Dietrich, Michael Schulze, Sebastian Zug, and Jörg Kaiser. 2010. Visualization of robot's awareness and perception. In *Proceedings of the First International Workshop on Digital Engineering*, ACM, Magdeburg Germany, 38–44. DOI:https://doi.org/10.1145/1837154.1837160

[7] Brian Duffy. 2003. Anthropomorphism and the social robot. *Robotics and Autonomous Systems* 42, (March 2003), 177–190. DOI:https://doi.org/10.1016/S0921-8890(02)00374-3

[8] Heike Felzmann, Eduard Fosch-Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2019. Robots and Transparency: The Multiple Dimensions of Transparency in the Context of Robot Technologies. *IEEE Robotics & Automation Magazine* 26, 2 (April 2019), 71–78. DOI:https://doi.org/10.1109/mra.2019.2904644

[9] Kerstin Sophie Haring, C^|^eacute;line Mougenot, Fuminori Ono, and Katsumi Watanabe. 2014. Cultural Differences in Perception and Attitude towards Robots. *IJAE* 13, 3 (2014), 149–157. DOI:https://doi.org/10.5057/ijae.13.149

[10] Minae Kwon. 2018. Setting Accurate Expectations of Robot Capabilities. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '18), Association for Computing Machinery, New York, NY, USA, 303–304. DOI:https://doi.org/10.1145/3173386.3176913

[11] Minae Kwon, Sandy H. Huang, and Anca D. Dragan. 2018. Expressing Robot Incapability. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '18), Association for Computing Machinery, New York, NY, USA, 87–95. DOI:https://doi.org/10.1145/3171221.3171276

[12] Minae Kwon, Malte F. Jung, and Ross A. Knepper. 2016. Human expectations of social robots. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 463–464. DOI:https://doi.org/10.1109/HRI.2016.7451807

[13] Laurel Riek, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, and Peter Robinson. 2009. *How anthropomorphism affects empathy toward robots*. DOI:https://doi.org/10.1145/1514095.1514158

[14] Nicolas Spatola and Olga A. Wudarczyk. 2021. Ascribing emotions to robots: Explicit and implicit attribution of emotions and perceived robot anthropomorphism. *Computers in Human Behavior* 124, (November 2021), 106934. DOI:https://doi.org/10.1016/j.chb.2021.106934

[15] Bernd Carsten Stahl. 2021. Ethical Issues of AI. In *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies*, Bernd Carsten Stahl (ed.). Springer International Publishing, Cham, 35–53. DOI:https://doi.org/10.1007/978-3-030-69978-9_4

[16] Andreas Theodorou. 2016. ABOD3: A Graphical Visualization and Real-Time Debugging Tool for BOD Agents.

[17] Andreas Theodorou, Robert H. Wortham, and Joanna J. Bryson. 2017. Designing and implementing transparency for real time inspection of autonomous robots. *Connection Science* 29, 3 (July 2017), 230–241. DOI:https://doi.org/10.1080/09540091.2017.1310182

[18] Jelte van Waterschoot, Merijn Bruijnes, Jan Flokstra, Dennis Reidsma, Daniel Davison, Mariët Theune, and Dirk Heylen. 2018. Flipper 2.0: A Pragmatic Dialogue Engine for Embodied Conversational Agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents* (IVA '18), Association for Computing Machinery, New York, NY, USA, 43–50. DOI:https://doi.org/10.1145/3267851.3267882

[19] Jelte van Waterschoot and Mariët Theune. 2021. Evaluating Conversational Question Generation: CoffeeBot. (2021).

[20] Robert H Wortham and Vivienne E Rogers. The Muttering Robot: Improving Robot Transparency Though Vocalisation of Reactive Plan Execution.

[21] Robert H. Wortham, Andreas Theodorou, and Joanna J. Bryson. 2017. Improving robot transparency: Real-time visualisation of robot AI substantially improves understanding in naive observers. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, IEEE, Lisbon, 1424–1431. DOI:https://doi.org/10.1109/ROMAN.2017.8172491

[22] 2023. Anthropomorphism | Definition, Examples & History | Britannica. Retrieved June 21, 2023 from https://www.britannica.com/topic/anthropomorphism

APPENDIX

A Descriptive Statistics

Descriptive Statistics				
	Interface	Mean	Std. Deviation	N
noViz_INTAvg	TUI	2,756	,8762	9
	GUI	2,867	,8000	9
	Total	2,811	,8159	18
viz_INTAvg	TUI	2,911	,8373	9
	GUI	3,044	,8531	9
	Total	2,978	,8229	18
noViz_LKBAvg	TUI	3,489	,8838	9
	GUI	3,733	,7348	9
	Total	3,611	,7984	18
viz_LKBAvg	TUI	3,556	,7601	9
	GUI	3,644	,7667	9
	Total	3,600	,7420	18
noViz_understandibility	TUI	1,222	1,3017	9
	GUI	1,778	,6667	9
	Total	1,500	1,0432	18
viz_understandibility	TUI	2,556	,8457	9
	GUI	3,333	,7071	9
	Total	2,944	,8556	18

Table 2

B Briefing letter

We are excited to have you join us for an upcoming experiment focused on human-robot interaction. In this experiment, you will have the opportunity to interact with our talking social robot, CoffeeBot. The aim of this study is to explore how people engage with and respond to the robot's questions and follow-up questions.

During the session, you will have conversations with CoffeeBot. The robot will initiate the interaction by asking you questions, and based on your answers, it may ask follow-up questions to further dive into the topic. CoffeeBot has been programmed to engage in dialogue, and generate questions based on your inputs. It is not as much of a chat bot, as it focuses on question generation.

If any of the questions of the robot makes you uncomfortable, or you don't know what to say, just answer I don't know or a generic answer.

The purpose of this experiment is to assess how well people understand robots. We are particularly interested in understanding how the robot's behavior is understood, and which factors affect the understandability of the robot, which refers to the clarity and openness of the robot's communication, allowing you to understand its intentions and reasoning.