Toward automated interactions for pediatric inpatients using a social robot

Aimée Sousa Calepso
aimee.sousa@honda-ri.de
Honda Research Institute Europe
Frankfurt am Main, Germany

Valerie Behrwind
valerie.behrwind@charite.de
Department of Pediatric Hematology and Oncology, Charité – Universitätsmedizin Berlin
Berlin, Germany

Christine Knoll*
christine.knoll@charite.de
Institute of Medical Informatics, Charité – Universitätsmedizin Berlin
Berlin, Germany

Sarah B. Blakeslee†
sarah.blakeslee@charite.de
Department of Pediatric Hematology and Oncology, Charité – Universitätsmedizin Berlin
Berlin, Germany

Bernhard Sendhoff
bernhard.sendhoff@honda-ri.de
Honda Research Institute Europe
Frankfurt am Main, Germany

Anna Belardinelli
anna.belardinelli@honda-ri.de
Honda Research Institute Europe
Frankfurt am Main, Germany

Abstract

We report on the deployment of the Haru social robot in a pediatric oncology ward to reduce social isolation and detail the main challenges faced during the process. We started with a semi-scripted interaction to gain insights about potential technical and psychosocial issues with such sensitive population, before moving to more flexible and automated conversations. We placed the robot in patients' single rooms and tested simple dialogues, after which we interviewed caregivers and children about the interaction. Such a preliminary effort revealed practical challenges that would critically affect the fluidity of the interaction, even if related to low-level, basic infrastructure. We also report on the introduction of the robot in a new environment and lessons learned on the importance of bringing technological awareness to this process.

Keywords

social robots, conversational agents, health care

ACM Reference Format:
Aimée Sousa Calepso, Valerie Behrwind, Christine Knoll, Sarah B. Blakeslee, Bernhard Sendhoff, and Anna Belardinelli. 2026. Toward automated interactions for pediatric inpatients using a social robot. In *Proceedings of IEEE/ACM Human-Robot Interaction (HRI '26).* ACM, New York, NY, USA, 4 pages.

1 Introduction

The use of social robots to support children's well-being has been widely researched in the field of HRI [3, 27]. Common use cases vary from entertainment [12] to education [10], and companionship [1]. In health care use cases, social robots have been demonstrated to be a valuable source of support [4, 22, 27]. Robots have the advantage of not getting tired or lacking the attention necessary for this audience. They can also adapt to specific conditions, such as when interacting with children who have special needs [8, 14, 25]. However, some aspects of the interaction between robots and humans still lack a natural flow, e.g., regarding turn-taking [24] or non-verbal cues [5]. Especially with children, conversations can be challenging because their ability to express themselves and to converse following social norms is still developing [18]. Recent advances in Large Language Models (LLMs) have begun to address these limitations [e.g., 15]. LLM-driven dialogue systems enable more fluid, context-aware conversations and allow robots to engage in real-time interaction without relying solely on predefined scripts.

In this paper, we describe our experience deploying the social robot Haru [6] in a real hospital setting and the challenges we faced and led to the need of transitioning from scripted dialogues to more automated interactions. We report on the robot's deployment process. We also offer reflections aimed at broadening the discussion on how social robots and LLMs should be integrated and evaluated in a real-world setting.

2 Background

Achieving a more natural interaction between humans and robots is usually done through different components. Social robots, more specifically in use cases involving children, should have a pleasant and friendly appearance. Moreover, the way they communicate – verbally or non-verbally – also needs to be tailored to this target audience.

2.1 Social Robots

Social robots have been deployed across a wide range of use cases and are particularly valuable when human presence or attention is limited. This is especially relevant in pediatric health care settings, where compromised health conditions can isolate children from their typical social environments. In such situations, social robots may offer an alternative source of interaction that helps mitigate this loss. As examples, PLEO [17], Arash [16], and Nao [2] have been deployed in hospital settings to assist children. These robots can not only serve as companions but also facilitate communication between patients and staff, motivate children to establish a routine,

*Also with Hasso Plattner Institute, Digital Health - University of Potsdam.
†Also with Charité Competence Center of Traditional and Integrative Medicine, Charité – Universitätsmedizin Berlin.

HRI '26, Edinburgh, UK
2026.

and provide support in educational contexts. For this reason, involving stakeholders from different areas is important to understand the role that a social robot can play in health care support [13].

2.2 Conversational Agents

In health care contexts, conversational agents show great promise, yet experimental evidence of their benefit is still being assessed clinically [11]. Conversational Agents (CAs) differ from text-based chatbots. They require more than Natural Language Understanding (NLU) and Natural Language Generation (NLG). A verbal dialog between machines and humans also requires Automatic Speech Recognition (ASR) and Text-To-Speech (TTS) capabilities to acquire spoken utterances and utter replies. Embodied agents further benefit from aligned non-verbal behavior; therefore, social cues such as gaze and lip-sync, as well as motor control, are also required [26]. To orchestrate all these different components, a Dialogue Manager (DM) is usually employed. It can be based on behavior trees, finite state machines, or, as more recently explored, LLMs [21].

Still, background noise, different language accents, and the usage of facial masks are a few of the factors that can impair the flow of dialogues. In addition, generating the appropriate response also requires context and awareness of the system. It does not only require appropriate semantics, but also the timing of when it happens [24]. Because dialogues are not always linear, interruptions should also be handled by the system. Since these steps require context, generating a dialogue that feels natural, is appropriate for the target audience, and has adequate latency is not a trivial task [9].

2.3 Generative AI and LLMs

Recent advances in Generative AI (GenAI) models have been used in all aspects of dialogue systems. In interactions with children, for example, research has investigated the use of AI-powered chatbots to support communication between health care professionals, patients, and their relatives [23]. Generative models can be applied in various ways. Voice synthesis and emotional tone adaptation are handled by TTS models, which can clone voices and modulate prosody. ASR models like Whisper [19] transcribe a speaker's speech into text. LLMs then interpret the transcribed text and generate appropriate, contextually relevant responses. This response generation is a particularly significant advance. While scripted dialogues often result in less engaging interactions, LLMs enable dynamic, automated conversations. However, this capability comes with a major challenge: LLMs are prone to hallucinations, can be manipulative or produce coherent but factually incorrect or contextually inappropriate outputs in unpredictable ways, which is particularly critical in health care contexts [20]. Future research must further investigate methods to predict and prevent such errors. Therefore, especially when engaging with vulnerable audiences, it is crucial to implement robust safeguards to ensure these technologies are used safely.

Still LLMs, can also help in alleviating ASR failures, such as correctly interpreting wrongly transcribed utterances. Finally, they can also automatically generate responses to the speaker. This is where large-language models enable the automatic creation of scripts within the context of the conversation. Scripted dialogues result in a less engaging interaction, so having a way to automate this step efficiently represents a significant advance in this field. At the same time they bring a big advantage, predicting the output of a generated dialog line is not trivial. LLMs are prone to hallucinations, meaning generated responses that are out of context in an unpredictable way. Current research still needs to further investigate methods to predict and prevent such mistakes. Therefore, in the context of a vulnerable audience, it is crucial to ensure LLMs are used safely.

3 Case study

We investigate the deployment of the social robot Haru in a highly sensitive health care environment. We worked with health care professionals and social scientists and developed a use case with patients in the pediatric oncology and stem cell transplant unit at the Charité Hospital in Berlin. We tested a simple interaction with a few patients with one parent/caregiver attending. We selected the pediatric oncology ward as the deployment site due to the availability of staff and hospital infrastructure. This setting also allowed for prolonged interactions with patients, who typically undergo treatment for several weeks, and remain isolated from their normal routine. While these patients can benefit the most from technological solutions to alleviate their social isolation, they are also very vulnerable users. With this use case, we hoped to uncover the challenges of introducing a social robot in such a sensitive environment, considering not only interactions with patients but also acceptance from staff, parents, and family members. In this first step, we opted to implement a semi-scripted dialogue to investigate the impact of Haru's embodiment and emotional expressions on such patients, to guarantee the appropriateness of the interaction.

3.1 Haru's introduction

The interaction we developed was simplified, focusing on introducing the robot in the pediatric ward. We started by interviewing the hospital staff to understand their needs and expectations. We collected data from social workers, nurses, doctors, and psychologists through semi-structured interviews. Their feedback was positive regarding having a social robot as a companion for children patients. One of the concerns raised was privacy and data protection, and the importance of ensuring the patient's safety. After reviewing the interview results, we developed an introduction to the robot for staff, parents, and patients. We distributed flyers to the parents showing pictures of Haru and explaining how it would interact with the patients. This step was important for setting expectations and preparing all parties involved.

3.2 Interaction

To tailor the interaction to our audience, we started by scripting dialogue and creating age-based branches, featuring jokes and riddles. We had one branch for younger children (up to 10 years old) and another one for older ones. Considering our target group, the interaction was designed to be relatively brief and shallow, to engage children of different ages while still providing a glimpse of the possible scenarios with such a robot. To provide a realistic impression and collect honest feedback from patients and their caregivers, the interaction was designed to be autonomous (no Wizard-of-Oz) but follow a pre-determined flow. Based on the patient's response, the dialog automatically proceeded to the intended branch.

Haru's response was conveyed through emojis displayed on the eyes, vocalizations, and neck and eye movements [6, 7]. Both verbal and motor reactions were aligned with the recognized intent of the patient's input. After asking about name and age, all children would go through some questions-and-answers-based chit-chat (e.g., *"What's your favorite season?"*). Depending on the branch, Haru then offered different entertainment options. Children aged 10 and above were offered scientific facts and/or riddles to guess. Children up to 10 could choose to listen to a short story. Finally, both dialog branches concluded in the same way: greeting the patient by name and asking if they had fun. The whole conversation was carried out in German. In total, we observed Haru interact with 8 patients aged 5 to 17 years old, and each interaction lasted about 15 minutes.

In terms of logistics, making sure the robot could be easily relocated to each of the rooms was important, especially while assuring that the connectivity with the server computer was not hampered. Within our immunocompromised patient group, we also ensured that the equipment was properly sanitized after each round of interaction.

3.3 Technical implementation

To perform speech recognition, we used the Whisper model. For entity recognition (e.g., favorite season, color, superpower,..), we structured the dialogue flow so that the system could rely on appropriate replies considering the recognized entity, or on generic reactions whenever an expected entity could not be made out throught the ASR. This avoided interrupting the conversation flow. For critical branching points (name, age, yes/no to offered service) questions would be repeated, in case the answer was not correctly acquired. For example, when asking the patient's age, the script expected a number as an answer and handled it accordingly. For riddles, we stored the correct responses in advance and compared the patient's answer to the predefined solution. Haru's voice was created based on a teenage voice model, to have an appropriate and genderless voice, fitting its character.

3.4 Technical issues

Regarding technical issues, we experienced problems with the network connection between the robot and the computer running the dialogue server. The delay particularly affected the ASR system's response time, which disrupted the natural flow of the interaction. Specifically, the slow processing of spoken input led to interruptions in turn-taking, making the system appear less responsive. These connectivity issues directly impacted the overall user experience, highlighting that introducing additional components also requires a reliable, low-latency communication infrastructure. This becomes even more important as the number of components required to create these dialogues increases, including the use of LLM models.

Besides the delays, we also noticed that the ASR model sometimes did not parse numbers consistently: it would transcribe them as numerals or as words. This problem would then affect the choice of the branch of the dialogue.

Because this first experience provided insights across the different stages of the deployment, we decided to address the main issues identified before proceeding to more complex interactions.

4 Open questions

While ethnographic observations and interview results are still being analyzed, on the basis of this experience, we could reflect on several questions. Feedback from teenage patients suggested that the interaction was rather mechanistic and for such older audience the overall experience was less engaging than for smaller children. While the population in pediatric wards span a 0-18 range, it needs to be considered that an engaging interaction should factor in age-appropriate modifications. In this LLMs could great improve the conversation. Still, other issues have to be tackled [20].

4.1 Patient's safety

First of all, how can we ensure that LLMs are safe to interact with children, particularly when dealing with vulnerable patients? This includes concerns re be addressed in the integration of LLMs in dialogues is avoiding sensitive topics, and making sure hallucinations can be prevented. This may vary according to different cultures, but conversations about illness and death, for instance, would need to be prevented at all costs, especially depending on the state of the patient. The present interaction was conducted only by voice (through a microphone), to gain trust from the ward personnel and parents, no camera was used, utterances were transcribed but not recorded, and all NLP processing happened locally. Still, this limited the responsiveness of the robot to the child reactions. So far, we can use a few strategies to prevent that certain topics are discussed by LLMs, such as prompting the model to avoid them at the start of each dialogue. There is no guarantee, so far, that this is enough to completely prevent inappropriate behavior, and further research needs to investigate alternative solutions. The strategies to prevent unwanted topics might also be useful to tackle hallucinations that might appear during interactions. We acknowledge that children might even find certain out-of-the-context answers funny, and answer questions with creative answers themselves, so there is a room for flexibility, as long as the patient's safety is ensured. Another approach that could provide more safety is implementing a list of key words that are banned from the conversation, filtering every dialog generated line according to it.

4.2 Data protection

In this respect, another important point is how to ensure that stakeholders understand that the data collected and analyzed during these interactions is protected and handled appropriately. In such situations, the parents or guardians of the patients need to consent to their participation in the studies. Moreover, the ethical board of hospitals also needs to be sure that no sensitive information is being used in a commercial or malicious way. In particular, questions regarding the storage and processing of audio and video recordings, as well as the handling of transcribed data, become relevant. The trade-off between data collection and user safety warrants further discussion to ensure privacy while still enabling research advances. The acceptance of data collection varies across cultures, potentially introducing bias into future research. We argue that, as a starting point, future research needs to raise awareness and provide information about these topics to ensure that experiments remain safe and accepted.

5 Conclusion

We reported on the first impressions of deploying Haru in a pediatric oncology ward. We implemented an interaction based on a semi-scripted dialogue, and tested it with 8 patients. Our experience brought us awareness of the potential issues that can arise when integrating GenAI models in multiple conversational AI components. While patients enjoyed the interaction, especially regarding the embodiment, the dialogues were still superficial, and the technical issues we faced made the turn-taking difficult at times. Expressions and non-verbal cues could only be produced after or before the verbal response, thus replies and non-verbal reactions were a bit disconnected. The interaction would be more engaging if the robot could be sensitive to the partner's gaze, face/voice expression. While this is technically feasible, it needs to be mediated with hospitals' regulations about patients' privacy. This use case provided us insights into the acceptability of social robots in such settings and prepared us to consider the next steps into more dynamic dialogues, while preserving the user's well-being and safety.

6 Acknowledgments

We would like to thank Randy Gomez and Eric Nichols from Honda Research Institute Japan, as well as Levko Ivanchuk for all the support provided to deploy Haru. We also would like to thank the members of the Service Robotics Lab from the Universidad Pablo de Olavide in Sevilla, Spain, for their support and insights in this project.

References

[1] Patrícia Alves-Oliveira, Elin A Björling, Patriya Wiesmann, Heba Dwikat, Simran Bhatia, Kai Mihata, and Maya Cakmak. 2022. Robots for connection: A co-design study with adolescents. In *Proceedings of the International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 578–583.

[2] Paul Baxter, Tony Belpaeme, Lola Canamero, Piero Cosi, Yiannis Demiris, Valentin Enescu, A Hiolle, I Kruijff-Korbayova, R Looije, M Nalin, et al. 2011. Long-term human-robot interaction with young users. In *Human-Robot Interaction Conference (robots with children workshop)*, Vol. 80. IEEE/ACM.

[3] Tony Belpaeme, Paul Baxter, Joachim De Greeff, James Kennedy, Robin Read, Rosemarijn Looije, Mark Neerincx, Ilaria Baroni, and Mattia Coti Zelati. 2013. Child-robot interaction: Perspectives and challenges. In *International conference on social robotics*. Springer, 452–459.

[4] Carlos A Cifuentes, Maria J Pinto, Nathalia Céspedes, and Marcela Múnera. 2020. Social robots in therapy and care. *Current Robotics Reports* 1, 3 (2020), 59–74.

[5] Alessio Galatolo and Katie Winkle. 2025. Simultaneous text and gesture generation for social robots with small language models. *Frontiers in Robotics and AI* 12 (2025), 1581024.

[6] Randy Gomez, Deborah Szapiro, Kerl Galindo, and Keisuke Nakamura. 2018. Haru: Hardware design of an experimental tabletop robot assistant. In *Proceedings of the ACM/IEEE international conference on human-robot interaction*. 233–240.

[7] Randy Gomez, Deborah Szapiro, Luis Merino, and Keisuke Nakamura. 2020. A holistic approach in designing tabletop robot's expressivity. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1970–1976.

[8] Bibi Huskens, Rianne Verschuur, Jan Gillesen, Robert Didden, and Emilia Barakova. 2013. Promoting question-asking in school-aged children with autism spectrum disorders: Effectiveness of a robot intervention compared to a human-trainer intervention. *Developmental neurorehabilitation* 16, 5 (2013), 345–356.

[9] Bahar Irfan, Sanna Kuoppamäki, Aida Hosseini, and Gabriel Skantze. 2025. Between reality and delusion: challenges of applying large language models to companion robots for open-domain dialogues with older adults. *Autonomous Robots* 49, 1 (2025), 9.

[10] Takayuki Kanda, Michihiro Shimada, and Satoshi Koizumi. 2012. Children learning with a social robot. In *Proceedings of the ACM/IEEE international conference on Human-Robot Interaction*. 351–358.

[11] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie YS Lau, et al. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association* 25, 9 (2018), 1248–1258.

[12] Iolanda Leite, Marissa McCoy, Monika Lohani, Daniel Ullman, Nicole Salomons, Charlene Stokes, Susan Rivers, and Brian Scassellati. 2015. Emotional storytelling in the classroom: Individual versus group interaction between children and robots. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*. 75–82.

[13] Leigh Levinson, Gloria Alvarez-Benito, J Gabriel Amores, Deborah Szapiro, Randy Gomez, and Selma Šabanović. 2025. Haru in the Care Network: Stakeholder Perspectives on Privacy with Social Robots in Pediatrics. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–28.

[14] Flavia Marino, Paola Chilà, Stefania Trusso Sfrazzetto, Cristina Carrozza, Ilaria Crimi, Chiara Failla, Mario Busà, Giuseppe Bernava, Gennaro Tartarisco, David Vagni, et al. 2020. Outcomes of a robot-assisted social-emotional understanding intervention for young children with autism spectrum disorders. *Journal of autism and developmental disorders* 50, 6 (2020), 1973–1987.

[15] Mauliana Mauliana, Ashita Ashok, Daniela Czernochowski, and Karsten Berns. 2025. Exploring LLM-powered multi-session human-robot interactions with university students. *Frontiers in Robotics and AI* 12 (2025), 1585589.

[16] Ali Meghdari, Azadeh Shariati, Minoo Alemi, Gholamreza R Vossoughi, Abdollah Eydi, Ehsan Ahmadi, Behrad Mozafari, Ali Amoozandeh Nobaveh, and Reza Tahami. 2018. Arash: A social robot buddy to support children with cancer in a hospital environment. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 232, 6 (2018), 605–618.

[17] Clara J Moerman and Rianne ML Jansens. 2021. Using social robot PLEO to enhance the well-being of hospitalised children. *Journal of Child Health Care* 25, 3 (2021), 412–426.

[18] Barbara A Pan and Catherine E Snow. 2016. The development of conversational and discourse skills. In *The development of language*. Psychology Press, 229–249.

[19] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*. PMLR, 28492–28518.

[20] Robert Ranisch and Joschka Haltaufderheide. 2025. Rapid integration of llms in healthcare raises ethical concerns: an investigation into deceptive patterns in social robots. *Digital Society* 4, 1 (2025), 7.

[21] Merle M Reimann, Florian A Kunneman, Catharine Oertel, and Koen V Hindriks. 2024. A survey on dialogue management in human-robot interaction. *ACM Transactions on Human-Robot Interaction* 13, 2 (2024), 1–22.

[22] Laurel D Riek. 2017. Healthcare robotics. *Commun. ACM* 60, 11 (2017), 68–78.

[23] Woosuk Seo, Young-Ho Kim, Ji Eun Kim, Megan Tao Fan, Mark S. Ackerman, Sung Won Choi, and Sun Young Park. 2025. Enhancing Pediatric Communication: The Role of an AI-Driven Chatbot in Facilitating Child-Parent-Provider Interaction. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1079, 16 pages. doi:10.1145/3706598.3713134

[24] Gabriel Skantze. 2021. Turn-taking in conversational systems and human-robot interaction: a review. *Computer Speech & Language* 67 (2021), 101178.

[25] Adriana Tapus, Andreea Peca, Amir Aly, Cristina Pop, Lavinia Jisa, Sebastian Pintea, Alina S Rusu, and Daniel O David. 2012. Children with autism social engagement in interaction with Nao, an imitative robot: A series of single case experiments. *Interaction studies* 13, 3 (2012), 315–347.

[26] Jacqueline Urakami and Katie Seaborn. 2023. Nonverbal cues in human–robot interaction: A communication studies perspective. *ACM Transactions on Human-Robot Interaction* 12, 2 (2023), 1–21.

[27] Xinghua Wang, Yu Wang, Zhongmei Han, Ziyu Duan, Zaipeng Zhang, and Taghreed Ali Alsudais. 2025. Social robots for child development: research hotspots, topic modeling, and collaborations. *Humanities and Social Sciences Communications* 12, 1 (2025), 1–15.