Exploring Human-Robot Interaction for University Open Days with Dialogue-Driven AI Assistants Using Speech Processing and Large Language Models

Margot M. E. Neggers, Edirlei Soares de Lima(✉), and Jonathan van Santen

Academy for AI, Games and Media, Breda University of Applied Sciences, Breda, The Netherlands
{neggers.m,soaresdelima.e,234275}@buas.nl

Abstract. This study explores the integration of Large Language Models (LLMs) and speech-processing technologies in humanoid robots to support during university open days. We implemented an AI-driven assistant on the Pepper and NAO robot to provide information to open day visitors. The system employs real-time speech recognition using OpenAI's Whisper, dialogue validation with GPT-4, and a Retrieval-Augmented Generation (RAG) approach to ensure accurate and context-aware responses. An exploratory observational study revealed positive user engagement, with many participants expressing surprise at the conversational abilities of the robot. However, challenges such as background noise, response latency, and user distrust were also observed. These findings suggest that LLM-powered robotic assistants can enhance visitor experiences in educational settings, suggesting areas for improvement in responsiveness, non-verbal interaction, and trust-building.

Keywords: Human-Robot Interaction · Observational Study · Dialogue-Driven Interaction · Robot Dialogue · Large Language Models

1 Introduction

Recent advances in Natural Language Processing (NLP) [28] have enabled significant progress in Human-Robot Interaction (HRI) [12], particularly through the integration of technologies such as large language models (LLMs), speech-to-text (STT), and text-to-speech (TTS) systems. These developments make it possible to support natural, context-aware dialogues between humans and robots, expanding their applicability in domains such as education [6], healthcare [7,26], and visitor assistance [10].

In this paper, we explore the potential of combining these technologies to create embodied conversational agents in university open days. These events

H. Degen and S. Ntoa (Eds.): HCII 2026, LNAI 16746, pp. 502–515, 2026.
https://doi.org/10.1007/978-3-032-31048-4_31

are key opportunities for institutions to inform and engage prospective students [21], but high visitor volume and limited staff availability can make it difficult to offer consistent and personalized guidance. To address this challenge, we implemented Artificial Intelligence (AI) assistants on the Pepper and NAO robots (Softbank Robotics) [19], designed to assist prospective students visiting Breda University of Applied Sciences (BUas). The assistants are capable of understanding natural language queries, retrieving relevant information from a structured knowledge base, and responding verbally in real time. This is achieved through an architecture that integrates an LLM for natural language understanding and generation, along with speech processing and context management strategies to ensure coherent interaction in dynamic environments.

The objectives of this study are twofold. First, we aim to implement a robust system architecture for real-time human-robot interaction, exploring the generative capabilities of LLMs to enable natural dialogue. By integrating speech processing to handle environmental noise and overlapping speech, the system is designed to handle the complexity of a real-world open day environment. Second, we conduct an exploratory observational study to assess the initial reactions and engagement of prospective students during interactions with the AI assistants at the open day.

The paper is organized as follows. Section 2 reviews related work. Section 3 details the technical implementation of the system. Section 4 presents the user study and its results. Section 5 discusses the findings, and Sect. 6 concludes with final remarks and recommendations.

2 Related Work

Robots have been explored in a variety of roles during university open days, including registration, navigation, and information delivery. For instance, Bellotto et al. [1] developed a robot capable of automatically recognizing and registering visitors, while Nguyen et al. [17] implemented a guiding robot to help attendees locate specific rooms or presentations. More recent work has focused on enabling robots to provide spoken information to visitors through conversational interfaces [23]. Like Shanthakumar et al. [23], in this presented work we provide visitors with information. However, where their focus is on comparing the functionalities of different LLMs, our aim is to develop a complete architecture and perform initial an user test.

Providing information through natural interaction is a common goal in many human-robot interaction studies. One example is the work by Blair and Foster [2], who developed a Pepper-based robotic assistant to offer guidance and support on a university campus. Their findings highlight the benefits of proactive robot behavior, such as initiating interactions, which improved user engagement. They identified several areas for improvement, including using signal processing to detect user intent, integrating Retrieval-Augmented Generation (RAG), and adopting LLMs to enhance dialogue naturalness and diversity. Our system addresses these challenges by incorporating RAG and LLMs to generate more informative and context-aware responses.

Recent studies have also explored the broader potential of LLM-powered social robots. Kim et al. [8] emphasized how LLMs enhance the expressive capabilities of robots while also increasing users' expectations regarding their conversational and social competence. Onorati et al. [18] presented a system in which speech processing, an LLM, and TTS were integrated to create a responsive conversational agent, which was positively evaluated in user studies. However, they reported limitations related to cloud-based LLM latency, speech recognition performance in noisy conditions, and LLM hallucinations. In our system, we address these issues through local speech processing, dialogue validation, and knowledge-grounded response generation using RAG.

3 Technical Implementation

3.1 System Overview

To support real-time natural language interaction in the context of university open days, we developed a modular architecture for embodied AI assistants that integrates LLMs, speech processing, and robot control. The system is designed to operate across two main components: the BUas AI Assistant Platform and the Robot Interaction Application, each responsible for distinct stages of the interaction pipeline. An overview of the full architecture is shown in Fig. 1.

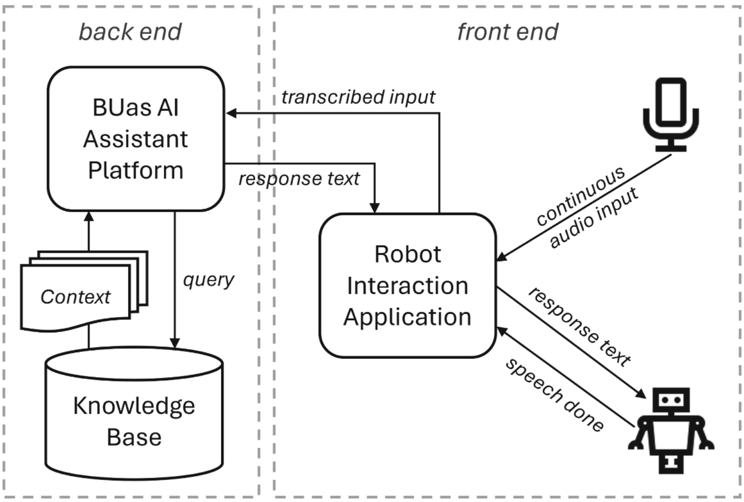

Fig. 1. Architecture of the proposed system.

The BUas AI Assistant Platform is a web-based system that manages chatbot interactions by defining the system prompt, providing access to a structured knowledge base, and processing user queries. The platform supports both open and proprietary LLMs and employs a RAG approach [11] using LangChain[1] to enhance response accuracy. The chatbot's responses are generated based on the conversation history and additional contextual information retrieved from the knowledge base.

Speech processing and dialogue validation are handled by the Robot Interaction Application, which runs locally on a laptop. This application captures audio input through a microphone and applies speaker diarization [20] and overlap detection [3] to identify the presence of multiple speakers. Once a single speaker is detected, the spoken input is transcribed using OpenAI's Whisper model.[2] The transcribed text is verified for coherence using an LLM (GPT-4o) before being forwarded to the AI assistant for response generation. This local execution of speech processing ensures reduced latency and improves the real-time performance of the system. Once a response is generated, the Robot Interaction Application transmits the corresponding response text to the robot, which then synthesizes and vocalizes the response using its onboard text-to-speech engine.

Communication between system components follows a structured protocol to ensure seamless data exchange. The Robot Interaction Application communicates with the BUas AI Assistant Platform via HTTP API calls, sending transcribed user input and receiving chatbot-generated responses. Once a response is received, the Robot Interaction Application forwards the corresponding text command to the robot via a TCP socket connection, instructing it to synthesize and vocalize the response. To prevent interference, the robot sends a "speech done" confirmation message back to the Robot Interaction Application once the speech output is completed. This mechanism ensures that speech recognition remains inactive while the robot is speaking, resuming only after the confirmation is received.

3.2 Speech Processing Pipeline

The speech processing pipeline is responsible for capturing user input, segmenting and transcribing speech, and preparing the transcribed text for further dialogue processing. Since the open day environment contains background noise and multiple conversations occurring simultaneously, the pipeline is designed to ensure robust speech recognition while minimizing erroneous inputs.

The pipeline begins with the audio capture process, where a microphone connected to the laptop continuously records incoming speech signals. The microphone is positioned close to the robot to maximize the clarity of the recorded voice while minimizing interference from surrounding noise. Incoming audio is processed in real-time, using a continuous recording approach that captures speech in small segments and stores them in a queue. The system applies an *energy threshold* of 1000 to detect when a speaker is actively talking,[3] ensuring

[1] https://www.langchain.com/langchain.

[2] https://github.com/openai/whisper.

[3] The energy threshold defines the minimum amplitude level required to distinguish speech from background noise.

that background noise does not trigger false detections. A phrase timeout of two seconds is used to determine when a speech segment has ended, marking the end of a spoken utterance before further processing.

To determine whether the recorded speech originates from a single speaker or multiple overlapping voices, the system employs speaker diarization and overlap detection. Speaker segmentation is performed using Pyannote's speaker diarization model v3.1,[4] which assigns different time segments to distinct speakers [4]. Meanwhile, Pyannote's segmentation model v3.0[5] detects overlapping speech segments. Both models are executed locally on a GPU to minimize processing delays. A speech segment is only passed to the transcription stage if exactly one speaker is detected and no overlap occurs, preventing the AI assistant from responding to unintended or mixed conversations.

For speech-to-text processing, the system uses OpenAI's whisper-large-v3 model running locally.[6] Whisper was chosen for its robustness in handling diverse accents, background noise, and spontaneous speech patterns, making it well-suited for an open day environment. The transcribed text is subsequently forwarded to the next stage of the pipeline, where it undergoes dialogue validation and processing.

3.3 Dialogue Validation and Response Generation

Once the speech-to-text processing is completed, the next stage involves validating the input and generating a response. This stage ensures that only relevant user input is processed and that appropriate responses are generated in alignment with the system's conversational goals. When generating a response, the system also retrieves accurate information from a structured knowledge base to ensure factual accuracy.

The validation step filters the transcription to ensure it is coherent, complete, and directed at the AI assistant. Since the system continuously captures audio from its environment, there is a risk that unintended speech or background conversations may be transcribed and processed as part of the user interaction. To prevent the system from responding to irrelevant or fragmented utterances, each transcribed message is evaluated by an LLM (GPT-4o). The validation mechanism verifies whether the input is a clear and complete statement or question directed at the AI assistant, taking into account the conversation history. This is achieved using a validation prompt (Prompt 1), which instructs the LLM to classify the input as either valid or invalid. In the prompt, $\{H\}$ represents the conversation history, $\{M\}$ refers to the user message, and $\{A\}$ stands for the assistant's name. If the model determines that the message is incoherent, fragmented, or likely part of an external conversation, it is discarded, ensuring that only relevant inputs proceed to the response generation step.

[4] https://huggingface.co/pyannote/speaker-diarization-3.1.
[5] https://huggingface.co/pyannote/speaker-segmentation.
[6] https://huggingface.co/openai/whisper-large-v3.

Prompt 1: *Conversation History: {H}*
Message to Evaluate: USER: {M}
Instructions: Determine if the above message appears to be a coherent and complete question or statement directed to {A}, based on the ongoing conversation. Review the message in relation to the context, checking if it logically aligns with a clear request or question for {A}. Respond with:
- [YES] if the message appears clear, coherent, and likely directed to {A}, even if it changes the subject.
- [NO] if the message is incomplete, fragmented, incoherent, or appears to contain mixed speech, suggesting it may not be a direct question or statement intended for {A}.
Respond only with [YES] or [NO]. No further explanation is needed.

Once validated, the transcribed input is processed for response generation using a combination of a predefined system prompt and retrieval-augmented generation. An LLM (GPT-4o) constructs responses based on the system prompt, which defines the assistant's role, personality, and interaction guidelines. Additionally, the RAG mechanism, implemented via LangChain, queries a structured knowledge base to retrieve relevant and factually accurate information, ensuring that responses are both informative and contextually appropriate. This response generation process is executed within the BUas AI Assistant Platform, which serves as the central module for managing chatbot interactions and conversations.

For our experiments, two distinct AI assistants were deployed, each operating with a specific system prompt that aligns with their objectives. NAO, functioning as a general assistant, provides information on all BUas programs, campus facilities, and general university information. In contrast, Pepper is designed to assist prospective students specifically interested in the Applied Data Science & Artificial Intelligence (ADS&AI) bachelor program. Despite their differences, both assistants utilize the same structured approach for response generation, combining the system prompt with retrieval-augmented generation. The complete system prompts for NAO and Pepper are provided in our supplementary material at: https://edirlei.com/buas/aioda-sup-material-1.pdf.

Once the response is generated, it is transmitted to the Robot Controlling Application, which is responsible for delivering the final output to the user. The response is sent as a text string and converted into speech using the robot's built-in text-to-speech engine. While NAO and Pepper utilize different hardware configurations, both robots rely on their onboard systems to synthesize speech and vocalize responses. Additionally, to maintain a lifelike presence during interactions, the robots continuously perform idle movements and facial tracking, ensuring that they remain visually engaged with users while delivering responses.

3.4 Ambient-Aware Reactions

To improve the overall user experience, the system incorporates a set of ambient-aware reactions designed to attract attention and respond to environmental conditions, making the robot appear more socially aware, proactive, and responsive. Three types of reactions are implemented: welcome reactions, conversation starters, and noise reactions.

Welcome reactions are activated at the beginning of an interaction session, either during system initialization or following a reset of the conversation context. These short messages are designed to provide context about the robot's role, encourage user engagement, and establish the assistant's persona. For example, Pepper may say, "*Hello and welcome! I'm Pepper, your friendly AI assistant for the open day*", while NAO might greet visitors with, "*Welcome! I'm Nao, an AI assistant robot here to help you explore all that Breda University of Applied Sciences has to offer.*"

Conversation starter reactions are used to proactively invite interaction when the robot is idle. These utterances are triggered at predefined intervals during periods of inactivity to attract the attention of nearby visitors. The content is varied and informal in tone, ranging from informative remarks about study programs to humorous comments that reflect the robot's persona. For instance, Pepper might say, "*Some call me a robot; I like to think of myself as a digital tour guide for data science and AI!*", while NAO could prompt with, "*Did you know our students work on real-world projects across various programs? Just ask if you'd like to know more.*"

Noise reactions are designed to address situations in which the robot detects overlapping speech or elevated background noise, both common in dynamic public environments. When the system identifies failed or invalid user input – typically due to multiple concurrent speakers or environmental interference – it marks the event as a likely result of ambient noise. At predefined intervals, the robot may respond with a noise reaction to acknowledge the surrounding activity. These utterances help maintain the robot's social presence and reassure visitors that temporary lapses in responsiveness can occur. For example, the assistant might say, "*Oops! Lots of voices make it tricky for me to focus. Could you say that one more time?*" or "*With all this excitement, my sensors get a little overloaded – don't hesitate to give me a nudge if I seem lost!*"

4 User Study

As previously mentioned, the proposed architecture was deployed on two humanoid robots: Pepper and NAO. During the open day, the Pepper robot was positioned indoors to assist visitors with information specific to the ADS&AI bachelor program, while the NAO robot was placed outdoors to provide general information about BUas and its full range of academic programs. Each robot was accompanied by a researcher who conducted the observational study described in the following subsections.

4.1 Methods

Participants. We observed a total of 22 interactions with the NAO robot and 12 with the Pepper robot, involving 47 participants in total. Group sizes ranged from one to three active participants per interaction. All participants engaged with the robot voluntarily, with minimal prompting from the researchers. Most were prospective students of the university, often accompanied by parents or friends. Five interactions involved current BUas students or staff members.

Materials and Set-up. Each robot was accompanied by a laptop positioned nearby, which served both as a status display and as the input device for audio processing. The screen displayed the university logo along with a simple status message, as illustrated in Fig. 2. The laptop's proximity to the robot was necessary due to its role in capturing audio via an external microphone.

(a) (b)

Fig. 2. Setup of the robots during the open day: (a) NAO, serving as a general assistant; and (b) Pepper, assisting prospective students in the Applied Data Science & AI program.

Both robots were stationed in relatively quiet areas to reduce background noise. Researchers remained at a distance to avoid influencing participant behavior or diverting attention away from the robot.

Procedure. Visitors who showed interest in the robots could approach and initiate interaction. In most cases, the robot initiated the conversation. Occasionally, the researcher provided minimal encouragement to begin the interaction, after which conversations progressed naturally.

During each session, a researcher documented the participant's background (e.g., prospective student, parent), the duration of the interaction, discussion topics, perceived user satisfaction, any technical issues, and additional observations. Conversations varied in length from 10 s to 6 min, with an average duration of approximately 90 s. After each interaction, participants were asked about their experience with the robot.

4.2 Results

Topics Discussed. The two robots served distinct roles during the open day. The NAO robot acted as a general assistant for all visitors, while the Pepper robot was dedicated to providing information about the ADS&AI bachelor program. This distinction was reflected in the topics covered during their interactions.

Approximately half of the conversations with NAO focused on general information about BUas study programs. Other frequently discussed topics included directions to buildings or presentations, and general information about the university. A smaller number of interactions addressed broader subjects such as student life or housing in Breda. In contrast, interactions with Pepper primarily revolved around the ADS&AI program. Participants asked questions related to the curriculum, such as the level of programming and mathematics, the structure of the first year, and examples of student projects. Some conversations also included more general university-related topics.

In both cases, occasional interactions included playful or exploratory dialogue. Visitors sometimes asked the robots about their names, abilities, or prompted them to perform actions like dancing or giving high-fives. Despite these unscripted requests, the assistants were able to engage meaningfully, showcasing the flexibility of the system. Overall, the knowledge base and system prompts were effective in supporting accurate and contextually appropriate responses. The AI assistants successfully answered the majority of questions posed by visitors.

Technical Issues. The most frequent technical challenge encountered was related to ambient noise. This was particularly noticeable for Pepper, which operated indoors in a highly active area. In some cases, the system struggled to isolate the speaker from background conversations. While many interactions proceeded without issue, in some instances participants needed to repeat their questions for the robot to process them correctly.

Speech recognition limitations were also observed with short or context-dependent responses such as "Yes" or "No". These inputs often lacked sufficient information for the system to produce a relevant response, especially when delivered too early, too quietly, or before the system was ready to process new input. Additionally, the system occasionally failed to recognize abbreviations for university programs.

Another issue was response latency. Although the system was optimized for real-time interaction, the time required to process input and generate a response using the LLM introduced occasional delays. In a few instances, this delay appeared to confuse or frustrate participants, consistent with prior findings in LLM-based HRI [18].

Overall User Satisfaction. Participants generally responded positively to their interactions with the robots. Of the 34 observed conversations, 23 included clear indications of positive user satisfaction. Very often participants even indicated to be pleasantly surprised with the capabilities and knowledge of the robot. In cases where feedback was neutral or negative, the cause was typically related to technical issues such as poor speech recognition caused by the environment noise or delayed responses. These problems occasionally disrupted the flow of the interaction.

A small number of visitors appeared hesitant or skeptical about interacting with the robots. In a few cases, participants preferred to speak with human assistants instead. However, some initially hesitant users chose to engage and were often positively surprised by the quality of the responses. Understanding the factors behind initial reluctance remains an interesting direction for future research.

5 Discussion

This study demonstrated the feasibility of deploying AI-powered embodied assistants to support prospective students during open days. By integrating LLMs with a robust speech-processing pipeline and proactive interaction strategies, the system facilitated fluid and informative conversations in a real-world setting.

One of the notable strengths of our implementation was the use of RAG to enhance the factual reliability of responses. Prior research has highlighted the risk of hallucinations when using LLMs in conversational agents [8,18]. By grounding responses in a structured knowledge base, our system effectively mitigated this risk, and no hallucinated or incorrect responses were observed. Participants responded positively to the interaction, often expressing surprise at the robot's conversational abilities and depth of knowledge. This reinforces findings from Kim et al. [8], who noted that LLMs significantly enrich robot dialogue and increase user engagement. Our results suggest that combining RAG with LLMs not only improves accuracy but also contributes to a more natural and satisfying user experience.

However, the use of LLMs and RAG introduced a common challenge: response latency. While the assistants were generally well received, some participants expressed confusion or frustration during longer pauses between user input and system response. Similar issues have been observed in prior work; for example, Onorati et al. [18] addressed this by adding filler utterances such as "*Let me think about this*" to maintain conversational flow. In future work, we aim to explore similar strategies to mitigate the impact of latency and further enhance the perceived responsiveness.

In addition to RAG, the speech processing pipeline contributed to the system's robustness in the noisy real-world environment of the open day. Techniques such as speaker diarization, overlap detection, and coherence validation helped reduce the likelihood of misinterpreting background conversations as user input. While challenges such as noise interference and delayed responses were occasionally observed, the overall stability and accuracy of the speech recognition system improved the user interaction.

Proactive behaviors, such as welcome and starter reactions, played a key role in attracting user attention and initiating conversations. These mechanisms align with prior findings suggesting that socially proactive robots are more likely to engage users [2]. Nevertheless, our study also revealed instances in which users hesitated to engage with the robot. Some visitors appeared uncertain about the assistant's ability to address their questions, indicating a potential mismatch between user expectations and perceived robot competence. This observation echoes earlier findings about the influence of task sensitivity and user trust in HRI [22].

A limitation of the current implementation is the need to manually reset the conversation context between interactions. Although the robot greets each new visitor with a welcome message, the system itself does not autonomously detect when one interaction has ended and another has begun. As a result, researchers occasionally intervened to reset the conversation history when a new participant approached. This limitation originates from the assistant's lack of situational awareness regarding multiple users. Future iterations could address this by integrating real-time engagement estimation techniques – such as gaze or depth-based tracking [27,30] – to help the system distinguish between separate interactions and trigger appropriate transitions more naturally [25]. Such mechanisms would also contribute to a more socially aware behavior model, enabling the robot to initiate context-appropriate reactions – such as greetings or farewells – based on the presence or departure of visitors.

Another important direction for future work concerns the integration of non-verbal behavior. While the robots currently perform idle movements and use face tracking to convey a sense of responsiveness – techniques known to enhance user engagement [5,9,16] – they do not yet incorporate meaningful gestures during speech. Prior studies have shown that LLM-driven conversations raise users' expectations for socially expressive behavior [8]. To address this, we are exploring the integration of gesture generation into the interaction pipeline. Preliminary tests have shown promising results, and future research will involve a comparative study evaluating user experiences with and without dynamic gesture capabilities. Furthermore, emerging approaches that enable LLMs to control robot gestures directly [29] present promising opportunities for achieving synchronized verbal and non-verbal communication.

While the exploratory nature of this study limits the generalizability of the findings, it provides a solid foundation for advancing socially intelligent robotic assistants in real-world settings. By demonstrating the feasibility and positive reception of LLM-driven interactions during open days, this work opens opportunities for more extensive studies related to the use of AI assistants in educational environments.

6 Conclusion

This study explored the use of AI-powered assistants integrated into humanoid robots to support prospective students during open days. Through the implementation of a modular architecture and an observational user study, we investigated how LLMs, combined with speech processing and ambient-aware behaviors, can enable real-time and natural interactions in dynamic environments.

The implementation demonstrated that LLMs and RAG can effectively mitigate hallucinations, while proactive behaviors and context-aware validation strategies enhance user engagement and interaction quality. Observational findings revealed that visitors generally responded positively to the robot's conversational capabilities. Nevertheless, the study also identified limitations such as response latency and sensitivity to background noise.

Overall, the findings highlight the feasibility and potential of deploying LLM-based embodied assistants in real-world scenarios, aligning with broader advances in conversational AI for educational contexts [15,24] and generative AI applications [13,14]. Future work will focus on enhancing interaction naturalness through gesture integration, automated visitor detection, and broader user studies to assess trust and usability across diverse audiences.

References

1. Bellotto, N., Rowland, S., Hu, H.: Lux-an interactive receptionist robot for university open days. In: 2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 1355–1360. IEEE (2008). https://doi.org/10.1109/AIM.2008.4601859
2. Blair, A., Foster, M.E.: Real-world evaluation of a university guidance and information robot. In: Ali, A.A., Cabibihan, J.J., Meskin, N., Rossi, S., Jiang, W., He, H., Ge, S.S. (eds.) Social Robotics, pp. 193–203. Springer Nature Singapore, Singapore (2024). https://doi.org/10.1007/978-981-99-8718-4_17
3. Bredin, H., Laurent, A.: End-to-end speaker segmentation for overlap-aware resegmentation. In: Proc. Interspeech 2021, pp. 3111–3115 (2021). https://doi.org/10.21437/Interspeech.2021-560
4. Bredin, H., et al.: Pyannote.audio: neural building blocks for speaker diarization. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7124–7128 (2020). https://doi.org/10.1109/ICASSP40776.2020.9052974
5. Cuijpers, R.H., Knops, M.A.M.H.: Motions of robots matter! The social effects of idle and meaningful motions. In: Tapus, A., André, E., Martin, J.C., Ferland, F., Ammi, M. (eds.) Social Robotics. Lecture Notes in Computer Science, vol. 9388, pp. 174–183. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25554-5_18
6. Hernandez-Cedeño, J., Ramírez-Benavides, K., Guerrero, L., Vega, A.: Nao as a copresenter in a robotics workshop - participant's feedback on the engagement level achieved with a robot in the classroom. In: Chen, J. (ed.) Advances in Human Factors in Robots and Unmanned Systems, pp. 143–152. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-319-94346-6_14
7. Khoo, W., et al.: Spill the tea: when robot conversation agents support well-being for older adults. In: Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, pp. 178–182. HRI '23, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3568294.3580067

8. Kim, C.Y., Lee, C.P., Mutlu, B.: Understanding large-language model (LLM)-powered human-robot interaction. In: Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, pp. 371–380. HRI '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3610977.3634966
9. Kompatsiari, K., Ciardo, F., De Tommaso, D., Wykowska, A.: Measuring engagement elicited by eye contact in human-robot interaction. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 6979–6985 (2019). https://doi.org/10.1109/IROS40897.2019.8967747
10. Leoste, J., Marmor, K., Hollstein, T., Hinkelmann, H., Leoste, L.B.: Enhancing university visitor satisfaction: a human-robot interaction study on the design and perception of a guiding robot assistant. In: Balogh, R., Obdržálek, D., Fislake, M. (eds.) Robotics in Education, pp. 223–234. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-67059-6_20
11. Lewis, P., et al.: Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. NIPS '20, Curran Associates Inc., Red Hook, NY, USA (2020)
12. de Lima, E.S., Feijó, B.: Artificial Intelligence in Human-Robot Interaction, pp. 187–199. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-319-96722-6_11
13. de Lima, E.S., Feijó, B., Casanova, M.A., Furtado, A.L.: ChatGeppetto - an AI-powered storyteller. In: Proceedings of the 22nd Brazilian Symposium on Games and Digital Entertainment, pp. 28–37. ACM (2024). https://doi.org/10.1145/3631085.3631302
14. de Lima, E.S., Neggers, M.M., Feijó, B., Casanova, M.A., Furtado, A.L.: An AI-powered approach to the semiotic reconstruction of narratives. Entertainment Comput. **52**, 100810 (2025). https://doi.org/10.1016/j.entcom.2024.100810
15. Mauliana, M., Ashok, A., Czernochowski, D., Berns, K.: Exploring LLM-powered multi-session human-robot interactions with university students. Front. Robot. AI **12** (2025). https://doi.org/10.3389/frobt.2025.1585589
16. Neggers, M.M., Ruijten, P.A., Cuijpers, R.H., IJsselsteijn, W.A.: Effect of robot gazing behavior on human comfort and robot predictability in navigation. In: 2022 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO), pp. 1–6 (2022). https://doi.org/10.1109/ARSO54254.2022.9802975
17. Nguyen, T.H., Tran, D.N., Vo, D.L., Mai, V.H., Dao, X.Q.: AI-powered university: design and deployment of robot assistant for smart universities. J. Adv. Inf. Technol. **13**(1) (2022). https://doi.org/10.12720/jait.13.1.78-84
18. Onorati, T., Castro-González, Á., Cruz del Valle, J., Díaz, P., et al.: Creating personalized verbal human-robot interactions using LLM with the robot mini. In: Proceedings of the 15th International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2023). Lecture Notes in Networks and Systems, vol. 835, pp. 148–159. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-48306-6_15
19. Pandey, A.K., Gelin, R.: A mass-produced sociable humanoid robot: pepper: the first machine of its kind. IEEE Robot. Autom. Mag. **25**(3), 40–48 (2018). https://doi.org/10.1109/MRA.2018.2833157
20. Park, T.J., Kanda, N., Dimitriadis, D., Han, K.J., Watanabe, S., Narayanan, S.: A review of speaker Diarization: recent advances with deep learning. Comput. Speech Lang. **72**, 101317 (2022). https://doi.org/10.1016/j.csl.2021.101317

21. Richardson, H.: University visit days. In: Cook, T., Rushton, B.S. (eds.) How to Recruit and Retain Higher Education Students: A Handbook of Good Practice, pp. 127–143. Routledge, London (2009). https://doi.org/10.4324/9780203877302
22. Rossi, A., Dautenhahn, K., Lee Koay, K., Walters, M.L.: How social robots influence people's trust in critical situations. In: 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 1020–1025 (2020). https://doi.org/10.1109/RO-MAN47096.2020.9223471
23. Shanthakumar, A.K., Fassihi-Tash, F., Lotfi, A., Bird, J.J.: Retrieval augmented large language model chatbots in higher education: a study on university open days. In: Advances in Computational Intelligence Systems. Advances in Intelligent Systems and Computing, vol. 1462, pp. 29–39. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-78857-4_3, uKCI 2024
24. Smit, K., Leewis, S., Almoustafa, H., Yildirim, K., Uymaz, T.: Enhancing educational dynamics integrating large language models with a social robot. In: Proceedings of the 2024 8th International Conference on Software and E-Business, pp. 87–94. ICSeB '24, Association for Computing Machinery, New York, NY, USA (2025). https://doi.org/10.1145/3715885.3715889
25. Sorrentino, A., Fiorini, L., Cavallo, F.: From the definition to the automatic assessment of engagement in human-robot interaction: a systematic review. Int. J. Soc. Robot. **16**(7), 1641–1663 (2024). https://doi.org/10.1007/s12369-024-01146-w
26. Su, Z., Sheng, W., Yang, G., Bishop, A., Carlson, B.: Adaptation of a robotic dialog system for medication reminder in elderly care. Smart Health **26**, 100346 (2022). https://doi.org/10.1016/j.smhl.2022.100346
27. Valikhujaev, Y.: MobileGaze: Pre-trained mobile nets for gaze-estimation (2024). https://doi.org/10.5281/zenodo.14257640
28. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems. vol. 30. Long Beach, CA, USA (2017)
29. Wu, Z., Shu, P., Li, Y., Li, Q., Liu, T., Li, X.: Robot Control via Natural Instructions Empowered by Large Language Model, pp. 437–457. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-66656-8_19
30. Yang, L., et al.: Depth anything V2 (2024). https://doi.org/10.48550/arXiv.2406.09414