Taylor & Francis
Taylor & Francis Group

Behaviour & Information Technology

ISSN: 0144-929X (Print) 1362-3001 (Online) Journal homepage: www.tandfonline.com/journals/tbit20

Generative AI-powered social robots in education: opportunities and challenges from a Delphi study

Gabriella Tisza , Panos Markopoulos , Sofia Serholt , Jauwairia Nasir , Omar Mubin , Adriana Tapus , Salvatore Anzalone , Koen V. Hindriks , Paul A. Vogt , Vasiliki Charisi , Emiel Krahmer , Mark A. Neerincx , Kaisa Väänänen, Daniela Conti , Jan de Wit & Emilia I. Barakova

To cite this article: Gabriella Tisza , Panos Markopoulos , Sofia Serholt , Jauwairia Nasir , Omar Mubin , Adriana Tapus , Salvatore Anzalone , Koen V. Hindriks , Paul A. Vogt , Vasiliki Charisi , Emiel Krahmer , Mark A. Neerincx , Kaisa Väänänen, Daniela Conti , Jan de Wit & Emilia I. Barakova (23 Dec 2025): Generative AI-powered social robots in education: opportunities and challenges from a Delphi study, Behaviour & Information Technology, DOI: 10.1080/0144929X.2025.2604060

To link to this article: https://doi.org/10.1080/0144929X.2025.2604060

Published online: 23 Dec 2025.

BEHAVIOUR & INFORMATION TECHNOLOGY
https://doi.org/10.1080/0144929X.2025.2604060

Generative AI-powered social robots in education: opportunities and challenges from a Delphi study

Gabriella Tisza[a], Panos Markopoulos[a], Sofia Serholt[b], Jauwairia Nasir[c], Omar Mubin[d], Adriana Tapus[e], Salvatore Anzalone[f], Koen V. Hindriks[g], Paul A. Vogt[h], Vasiliki Charisi[i], Emiel Krahmer[j], Mark A. Neerincx[k], Kaisa Väänänen[l], Daniela Conti[m], Jan de Wit[j] and Emilia I. Barakova[a]

[a]Department of Industrial Design, Eindhoven University of Technology, Eindhoven, Netherlands; [b]Department of Applied IT, University of Gothenburg, Gothenburg, Sweden; [c]Faculty of Applied Computer Science, Institute of Computer Science, Ausburg University, Ausburg, Germany; [d]School of Computing, Data and Mathematical Sciences, Western Sydney University, Sydney, Australia; [e]Autonomous Systems and Robotics Lab, Institut Polytechnique de Paris, Paris, France; [f]Université Paris 8, Saint-Denis, France; [g]Faculty of Science, Artificial Intelligence, Vrije Universiteit Amsterdam, Amsterdam, Netherlands; [h]Faculty of Science and Engineering, University of Groningen, Groningen, Netherlands; [i]Joint Research Centre, Unit of Digital Economy, European Commission, Brussels, Belgium; [j]Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands; [k]Human-Centered Computing, Delft University of Technology, Delft, Netherlands; [l]Faculty of Information Technology and Communication Sciences, Human-Centered Technology, Tampere University of Technology, Tampere, Finland; [m]Department of Humanities, University of Catania, Catania, Italy

ABSTRACT
The rise of Generative Artificial Intelligence (GenAI) and Large Language Models (LLMs) is accelerating the integration of social robots into education. These technologies enhance robots' abilities in natural language interaction, adaptive behaviour, and personalised learning support. To advance real-world implementation, it is essential to identify the main challenges and opportunities in this field. We conducted a two-round Delphi study with 16 experts in human-robot interaction and educational technology. In the first round, participants outlined opportunities, challenges, and potential robot roles expected in the short term (1 year) and medium term (5 years). Content analysis revealed 8 opportunities, 10 challenges and 10 roles. In the second round, experts ranked their importance and feasibility across both time horizons. The results show that the most critical opportunities and challenges are also the least feasible to achieve in practice. Conversely, the proposed roles of educational robots demonstrated alignment between importance and feasibility. Experts highlighted three promising roles for robots in the GenAI era: supporting teachers in boosting learner engagement, serving as conversational interfaces for students to access knowledge and assisting teachers in supporting disadvantaged learners. These findings provide a roadmap for prioritising feasible innovations in educational robotics.

ARTICLE HISTORY
Received 10 May 2025
Accepted 8 December 2025

KEYWORDS
Social robots; education; generative AI; Delphi study

1. Introduction

For over two decades, researchers have been investigating how learning can be supported by social robots capable of interacting with learners using human-like modalities such as speech, head and body movements, gaze, and facial expressions. Social robots emerge as a significant innovation in education, with applications spanning primary schools (Konijn et al. 2022) to higher education institutions (Rosenberg-Kima, Koren, and Gordon 2020). Social robots can serve as tutors or peer learners in areas such as language education (Vogt et al. 2017), music practice (Song, Tsiakas, et al. 2024), handwriting and calligraphy (Tozadore et al. 2022) or question-formulating skills (Huskens et al. 2013), bringing technology-driven learning experiences closer to how teachers have traditionally interacted with learners (Belpaeme, Kennedy, et al. 2018). They may be used alone or in conjunction with other devices such as tablets (e.g. Hei, Zhang, and Tapus 2024; Nasir et al. 2020; Vogt et al. 2019) or tangibles (e.g. Charisi et al. 2021; Davison et al. 2020) to deliver instructions, to motivate learners, and to provide them with feedback during learning.

Implementing social robots in education has traditionally been hindered by technological challenges. It has been notoriously difficult to design and develop social robots able to interact in a timely and socially appropriate manner, largely due to limitations in

CONTACT Gabriella Tisza gtisza.phd@gmail.com Department of Industrial Design, Eindhoven University of Technology, Eindhoven 5612AE, Netherlands
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

Artificial Intelligence (AI) technology (Belpaeme, Kennedy, et al. 2018; Honig and Oron-Gilad 2018; Serholt 2018; Serholt et al. 2020). The robots deployed in many of the empirical studies in this field have been tele-operated (Wizard-of-Oz set-ups) or could only work autonomously within narrowly defined boundaries regarding the interaction and the learning content (Barakova et al. 2015). The recent acceleration in Generative AI (GenAI) and foundation models, including Large Language Models (LLMs) for natural language understanding and generation (Vemprala et al. 2024), but also models for gaze, face, object recognition, and robot control bring the application of social robots much closer to educational practice. These developments give rise to questions regarding the feasibility and timeline for implementing LLM-powered robots in education, the opportunities they bring about and the challenges that need to be addressed to do so.

To explore how recent advances in GenAI present both opportunities and challenges for using social robots in education, we carried out a Delphi study (Strauss and Zeigler 1975) in which we invited a panel of experts in Human–Robot Interaction (HRI) and educational technology to discuss opportunities, challenges to overcome, and robot roles to be fulfilled for the application of LLM-powered robots in education within a 1-year and a 5-year timespan. Our study aims to provide the field with critical insights into the gaps limiting the implementation of social robots in education and to identify key priorities for future research. Furthermore, we aim to guide future researchers in the field and provide a resource for those working towards the practical implementation of social robots in education.

The following sections present a comprehensive review of related research on social robotics in education and especially the current state of research regarding GenAI-powered robots in education. We outline the rationale and methodology employed in this study and discuss the results in detail, offering a roadmap for addressing the identified challenges and opportunities.

2. Related work

2.1. Social robots in education

There has been substantial research interest in integrating social robots into formal learning environments. Compared to the more widely used screen-based technologies (e.g. tablets, smartboards), research has demonstrated that social robots can be more engaging for learners due to their physical presence, their ability to facilitate human-like social interaction (Brown, Kerwin, and Howard 2013; Mann et al. 2015; Van Lehn 2011) and tangible learning experiences with immediate feedback that enhance student motivation and interest for STEM education (Spolaôr and Benitti 2017) and arts/music education (Song, Tsiakas, et al. 2024). They can provide embodied social support through the learning process; e.g. a 3-week long deployment of a social robot in a classroom for children with ASD fostered a steady positive impact on their well-being (Lemaignan et al. 2024). They can guide learners, provide direct feedback (Park, Kim, and del Pobil 2011), and help create a safe, supportive learning environment, facilitating knowledge acquisition and enhancing engagement and motivation. Social robots can also facilitate adaptive and personalised interactions with learners, recognising emotional responses, applying personalised motivational strategies, and adapting to the preferences, requirements, and needs of each child (Obaid et al. 2018). By tailoring the robot's behaviour to align with the learner's motivational orientation, personalised tutoring by robots has been shown to enhance the efficiency of learning but also to increase engagement and interest, offering an adaptive and effective learning experience (Hei, Zhang, and Tapus 2023).

There is also growing evidence on how social robots can help improve learning outcomes. A recent meta-analysis on the use of social robots in language learning concluded that robots' socially supportive behaviours can have a positive effect on language learning achievement (Wang and Cheung 2025). A recent meta-analysis of 24 studies that examined the effects of educational robots on students' learning outcomes demonstrated a moderate but significant positive effect (Wang et al. 2023).

Robots can promote metacognition, helping learners tackle cognitively demanding tasks and achieve measurable learning gains (Ramachandran, Huang, and Scassellati 2017). Modelling robot's interactions after self-determination theory (van Minkelen et al. 2020) and evaluation apprehension theory (Song, Barakova, et al. 2024) have been shown to stimulate learners' intrinsic motivation, facilitating a sense of autonomy, competence, and relatedness, which was demonstrated to be effective when children learned a second language from a social robot (van Minkelen et al. 2020) and when children learned to cope with their chronic disease (Neerincx et al. 2019). For some tasks robots have been shown to surpass human tutors. Hei, Zhang, and Tapus (2024) compared users' help-seeking from a human tutor, a robot tutor, and a robot tutor, showing that the latter was more frequent, resulting in quicker assistance and better cognitive performance. To design the behaviours of the robot we can draw on psychology-based principles commonly applied in education to shape instructional strategies, motivate students, and foster productive learning environments. Hei, Zhang, and Tapus (2023) applied the Regulatory Focus Theory, which classifies motivation into two types: promotion-focused, driven by growth and achievement, and prevention-focused, driven by avoiding negative outcomes. They suggest that promotion-focused robots enhance learning efficiency for promotion-oriented participants, while prevention-focused robots increase learning interest for prevention-oriented participants. Song, Barakova, et al. (2024) report that learners perform better with a robot that did not provide evaluative comments. Song, Tsiakas, et al. (2024) learner had higher motivation and better performance when interacting with a robot that stimulates self-assessment, rather than a non-evaluative robot.

To different degrees, researchers consider robots as complementary to or as substitutes for teachers' work. Social robots may serve roles traditionally served by human teachers, but more common is the use of social robots as a peer or learning companion to the learner (Michaelis and Mutlu 2018; Song et al. 2021). Common motivations for such works are to overcome the limited availability of human teachers in different contexts, to personalise learning activities, to improve learning experiences through higher engagement of learners and/or to offer more opportunities to practice skills that need to be learned, to partake in co-curricular activities that support the curriculum, and/or to take over repetitive tasks that human teachers may find tedious.

However, introducing social robots in learning environments does not always produce the expected benefits for engagement and learning (Nasir et al. 2020; Riedmann, Schaper, and Lugrin 2024). Challenges such as technical limitations, imperfections in social behaviours (e.g. accents, nuanced feedback perceived as mismatched to learners' competence), and a constrained range of social interactions have occasionally diminished their effectiveness, as highlighted by Bravo Perucho and Alimardani (2023) and (Riedmann, Schaper, and Lugrin 2024). Social robots may even be counterproductive to learning goals because of untimely interventions, as suggested by Nasir, Bruno, and Dillenbourg (2024) and Ramachandran, Huang, and Scassellati (2017), or overly social behaviours (Kennedy, Baxter, and Belpaeme 2015; Yadollahi et al. 2018), which distract students from the learning process.

Multi-session studies of social robots report novelty effects, though these are not straightforward. Children have been found to have higher motivation during initial sessions with a robot compared to later ones (Kanda, Shimada, and Koizumi 2012), and social reactions towards robots were mixed with facial and verbal expressions declining after the initial exposure while the gaze and gestures towards the robot remained stable. Contextual factors might negatively influence a robot's effectiveness as a tutor (Kennedy, Baxter, and Belpaeme 2015). While some researchers have deployed educational robots to operate autonomously in a field context for sustained periods of time (e.g. Davison et al. 2020; Nasir et al. 2023, 2024; Ramachandran, Sebo, and Scassellati 2019; Serholt et al. 2017), the practical implementation of these robots is has not yet been demonstrated. According to Barakova et al.'s (2023) meta-review of research (i.e. review of the systematic and narrative review papers in the domain), social robots in education are still not reaching their full potential, and there seems to be a gap between practice and research considering how few social robots are being used in comparison to the research efforts undertaken in this area. Even in areas where the value of social robots is widely recognised, such as for children with ASD (Alnajjar et al. 2020), and where randomised controlled trials found that the interaction between children and teachers/parents improved due to the robot, while the interaction with the robot decreased (van den Berk-Smeekens et al. 2022; van Otterdijk et al. 2020), educators hesitate to adopt robots for fear of them disrupting teaching processes, increasing their workload, and that robots might replace interpersonal relationships. Policymakers, international organisations, and researchers have expressed concerns about the protection of children's rights when interacting with AI-based technologies, including social robots (Šabanović et al. 2023; Tolksdorf et al. 2021).

2.2. LLMs in education

The emergence of LLMs has opened the potential to address long-standing challenges through their remarkable capacity for reasoning and generation. LLMs can interpret ambiguous human language and process multimodal input, generating the desired output while reducing labour-intensive operations. LLMs are a breakthrough in handling language-based tasks such as producing coherent and contextually relevant text, answering questions, translating across languages, summarising, determining the sentiment expressed in text, extending to areas they have not been originally intended for, such as arithmetic, programming, following instructions, or in information retrieval and computer vision.

There has already been substantial interest in using LLMs as tutors for various subjects (Hu et al. 2025). Sonderegger (2022) proposed, but did not implement, a model connecting the different forms of cognitive engagement of the learner with the activities of the robot tutor, classifying the possibilities of generative language models. Other uses of LLMs in formal education include personalising teaching, creating lesson plans, summarising texts and highlighting main points, evaluating student works, or doing plagiarism checks (Caines et al. 2023; Kasneci et al. 2023). Dai et al. (2023) report an early exploration of using ChatGPT to provide students with feedback. They reported that it generated detailed, fluent, and coherent feedback that effectively summarised students' performance, aligning well with human instructors in assessing assignments and it provided process-oriented feedback, thereby aiding students in developing learning skills. Nichols, Gao, and Gomez (2020) showed how the reasoning ability of LLMs can enable chatbots to generate coherent stories for children, guiding the plot in the direction desired by the user. Deng et al. (2024) used multi-modal LLMs to assess social reciprocity in children with ASD.

These examples are only a small selection of a growing research field examining the various potentials of LLMs in education, which include tools for the teacher and the student, but also higher-level investigations regarding the current limitations of LLM technology for this application domain and the sociotechnical issues surrounding the introduction of LLMs in education.

2.3. LLM-powered robots

There have already been a few attempts to combine GenAI and LLM and social robots. Technical developments such as Pepper Chat which utilises Google Cloud speech-to-text functionality for speech-based dialogue, enable the efficient implementation of natural language dialogues, and a more natural and responsive communication than has so far been possible, enhancing the robustness and versatility of the robot in handling a wide range of input variations, and in different application domains (Billing, Rosén, and Lamb 2023). Wang et al. (2024) developed an LLM-based conversation system that dynamically generates expressive robot behaviour directly from the LLM during conversations. In a small-scale evaluation study, their system was found enjoyable, and LLMs could help recover from speech recognition errors. Kim, Lee, and Mutlu (2024) report a user study that compared LLM-powered robots against text- and voice-based agents in conversational tasks that included choosing, generating, executing, and negotiating. They found that the physical embodiment of LLM-powered robots may be preferred to virtual agents, particularly for connectedness and deliberation between user and robot, which is particularly relevant in educational settings. On the downside, LLM errors, including hallucinations and repetitions, were identified as threats for derailing conversations and hampering adoption and that the robot embodiment may induce anxiety and fall short of expectations in logical communication. Mahadevan et al. (2024) proposed a framework for generating and modifying expressive robot motion to accompany social interactions with humans, using LLMs to translate user language instructions to robot code, learning in-context with few-shot prompting. This approach is more flexible than rule-based programming and avoids the need for curated datasets, as in data-driven approaches.

The first published review of LLM-powered robots by Zhang et al. (2023) argues that the integration of LLM and robotics can radically transform HRI in terms of natural language understanding, the fluidity and intuition of human–robot collaborations, and the complexity and contextual embedding of the interaction. However, they also point out how this integration bares challenges with regards to safety, ethical considerations, and the genuine comprehension of context. Atuhurra (2024) reviewed 250 papers published in conferences focusing on HRI, aiming to understand how to safely and responsibly utilise LLM and Visual Language Models (VLMs) to develop social robots They recommend: (i) identifying and mitigating bias within LLMs (ii) developing secure, privacy preserving models that minimise data leakage and protect user information, (iii) model compression and efficiency to enable the deployment of LLMs on resource-constrained platforms (iv) leveraging LLMs to foster richer, natural HRI through integrating visual, auditory, and textual data, (v) establishing ethical guidelines for LLM-powered social robots, incorporating societal norms and human traits such as trust, politeness, personality, and gender considerations.

The emergence of LLMs may radically transform and advance the field of social robotics. Much of our understanding of social robot interaction has been gained in largely controlled experiments and very often through Wizard of Oz setups, where the robots are operated by a human, to evaluate different behaviours and roles for the robots. As researchers experiment with LLM-powered robots, these understandings may need to be re-examined, as LLM-powered robots with their enhanced conversational abilities seem to elicit heightened expectations for sophisticated non-verbal cues (Kim, Lee, and Mutlu 2024). Compared to screen-based interaction with LLMs, LLM-powered robots connect the LLM with the real world through the robots' embodiment and especially their sensing and actuation capabilities. Multi-modal LLMs can enhance the robot's nonverbal behaviours, which may be valuable in enhancing task interactions.

In an empirical investigation from a Design Justice perspective, Markelius (2024) examined ethical concerns regarding the combination of LLMs with social robots. These concerns include (i) potential emotional dependence and mental health harms, power divides reinforced through dominant language, and the combination of LLM output with non-verbal behaviour of the robot violating social norms and ethics; (ii) reproducing and exacerbating divides and harmful stereotypes; (iii) the risk of spreading misinformation; (iv) privacy attacks or data exploitation risks surrounding the combination of verbal utterances with audio-visual data collected by the robot.

2.4. LLM-powered robots in education

So far, there are relatively few attempts to use LLM-powered robots in education. Höhn, Nasir, Paikan, et al. (2024) and Höhn, Nasir, Tozadore, et al. (2024) explored the use of LLM-powered social robots and agents as social role-playing companions – an area that was previously challenging due to limitations in natural language understanding and generation. This approach has a broad range of potential applications, including supporting cognitive, physical, social, and emotional development in young children, aiding social skills training and second language learning for adults, and offering companionship to the elderly. Topsakal and Topsakal (2022) combined Augmented Reality, Voicebots, and ChatGPT for foreign language learning. Regarding special needs education, robots that can display facial expressions, such as Furhat Moxie, QTrobot, and EVA, Socibot, can be particularly relevant, though a systematic literature survey of related applications (Voultsiou et al. 2025) could not identify any work on LLM-powered robots for special education, although they did find relevant LLM applications for sign language recognition and the production of context-aware gestures for the deaf.

Allgeuer, Ali, and Wermter (2024) demonstrated a modular approach by which different sensing and actuation functions of the NICOL robot (speech recognition, speech generation, open-vocabulary object detection, human pose estimation, and gesture detection) can be combined flexibly with various LLMs, which would serve as a central text-based coordinating unit, to produce natural and appropriate HRI without explicit programming. Addlesee et al. (2024) demonstrated how LLMs could help recognise emotions in the dialogue utterances and trigger relevant facial expressions on the Furhat robot. Zhao et al. (2025) leverage LLMs for generating robotic actions based on multimodal sensory input, while Irpan et al. (2022) contextualise robot commands based on their perception of the environment, while the language model supplies high-level semantic knowledge about the task.

2.5. Towards the implementation of LLM-robots in education

Despite the wide research interest in using social robots for educational purposes, relatively few studies have examined how to implement them effectively in education. Belpaeme, Kennedy, et al. (2018) emphasise that educational robots should be designed to align closely with pedagogical objectives, ensuring that their interactions support the learning process and complement the teacher's role by providing social support, timely feedback, and encouragement. Smakman (2023) provides a list of guidelines to support implementing social robots in formal education in a morally justified way. Similarly, Lemaignan et al. (2021) proposed a set of guidelines that align the design, development, and use of AI applications, including robots for children, with children's rights as have been proposed by UNICEF.

Curriculum designers need to craft tasks and frameworks that facilitate the seamless integration of robots into the learning environment (Mubin et al. 2013). Van Lehn (2011) highlights the critical importance of customising a robot's behaviour and feedback to accommodate diverse learning styles and levels, which is critical for delivering personalised education. Both teachers and students should receive adequate training to be able to use robotic systems in the classroom effectively (Kennedy, Baxter, and Belpaeme 2015). According to Westlund et al. (2018) continuous, cyclical evaluation and adaptation of robot roles are essential to respond to the evolving educational demands and ongoing technological advancements, which includes appropriate IT infrastructure to maintain, troubleshoot, and ensure smooth functioning of the robots.

To implement social robots in formal education, it is crucial to involve not only children and teachers but all stakeholders. Smakman, Vogt, and Konijn (2021) investigated moral considerations for children and teachers, but also parents, robot developers, and governmental policymakers, and found that their priorities may diverge They all want to gain insights from data that robots can collect, however, for different reasons: teachers want data to improve teaching, policymakers to improve educational policies, the robotic industry wants data to help improve their products, and parents express their right to access all data recorded of their children (Smakman, Vogt, and Konijn 2021). A focus group study involving 77 teachers (Serholt and Barendregt 2016) into the ethical perspectives and social implications of introducing robots in the classroom, concluded that teachers were willing to consider robots as a teaching tool to facilitate children's 'robotic' literacy but were skeptical regarding autonomous classroom robots.

Earlier studies have indicated that social robots are expected to play a practical role in personalised tutoring, where robots can provide tailored instruction, adapting to individual students' needs and learning paces, thereby influencing students' motivation, comprehension, and learning outcomes (Hei, Zhang, and Tapus 2023, 2024; Prather et al. 2023; van Minkelen et al. 2020). Another opportunity lies in collaborative learning environments, where robots can facilitate teamwork, improve communication skills, and encourage peer interaction in problem-solving settings (Denis and Hubert 2001), and in dyad settings in open-ended computational learning settings (Nasir et al. 2023, 2024). It is expected that social robots will also be able to assist with classroom management, providing positive reinforcement, social cues, and emotional support, which can reduce the cognitive load on teachers and improve the classroom climate. It has been argued that robots could support inclusive education by assisting students with disabilities, offering personalised assistance, and enabling greater access to learning (Encarnação et al. 2017). Such expectations are largely formed before the emergence of LLM powered robots which calls for answering the following research question:

> RQ1. Considering advances in GenAI and LLMs, what are potential opportunities for near-term and mid-term practical applications of social robotics in education?

The practical use of social robots in educational settings falls considerably short of what the current state of research suggests would be possible (Barakova et al. 2023). One of the primary problems is the high cost and complexity of robot design, development, and maintenance, which limits access to schools and institutions with limited budgets (Belpaeme, Kennedy, et al. 2018). Additionally, current robots lack the capabilities needed to engage in meaningful, adaptive interactions with students over extended periods (Serholt 2018). With the arrival of GenAI and LLMs, social robots are coming closer to overcoming their main shortcoming: that they could not interact reliably using natural speech, especially regarding speech from children, second language speakers, or speakers with dialects (Alnajjar et al. 2021; Patel and Scharenborg 2024). Recent reports suggest that children's speech can now be sufficiently transcribed to support spoken child–robot interaction, especially when complemented by LLM for dialogue management, though the responsiveness of current cloud-based solutions remains a challenge (Janssens et al. 2024).

Scalability is another concern, as designing robots that can be easily integrated into various educational contexts, subject areas, and age groups is difficult (Westlund et al. 2018). Privacy, data security, and additional ethical concerns, particularly regarding the collection of student performance and behavioural data, also pose barriers to adoption, as schools must ensure compliance with regulations and safeguard the collected information about their pupils (Charisi et al. 2021; Serholt et al. 2017). A mixed-methods review of related research emphasised how several technical challenges need to be overcome to enable the use of robots in education: improving robot audio and video function, designing non-distractive gestures for robots, among others (Chou et al. 2023). A review of social robotics for education (Youssef et al. 2023) characterised two challenges that need to be addressed for robots to be used in education: technological and user-oriented challenges. Technological challenges relate to AI in general; for example, enhancing the cognitive capabilities on robots requires efficient signal processing, pattern recognition, decision making and response generation, but also developments in hardware. User-oriented challenges include designing learning materials and the curriculum defining the role of the human teacher and adapting robot behaviours to the learner, their age, and knowledge. Gentile et al. (2023) emphasise how the low level or even absence of adequate digital skills of the teachers presents a major barrier for the introduction of AI in the classroom, an observation that, by implication, also holds for LLM-powered social robots. Charpentier et al. (2022) highlighted the need for new didactics and pedagogies that explicitly address the opportunities and challenges brought by the presence of robots in classrooms. One way to overcome these challenges could be the explicit involvement of stakeholders in designing robot activities through collaborative approaches (Zou et al. 2024). Aiming for the successful implementation of social robots in education, the question arises:

> RQ2: How do recent advances in LLMs and GenAI specifically impact the key challenges for the widespread adoption of social robots in education in the near and mid-term future?

2.6. Robot roles in education

In order to derive knowledge that transcends individual robotic platforms, technical features, or social cues of social robots, researchers have been investigating various social roles for robots, e.g. the role of a tutor in learning language (Kennedy et al. 2016), geography and sustainability (Serholt et al. 2017), and mathematics (Ligthart et al. 2023; Smakman, Vogt, and Konijn 2021). Others have examined how robots can be cast in the role of a learner (Hood, Lemaignan, and Dillenbourg 2015; Pareto, Ekström, and Serholt 2022; Sandygulova et al. 2020; Yadollahi et al. 2018), a peer (Balkibekov et al. 2016; Nasir, Bruno, and Dillenbourg 2024; Norman et al. 2022; Park et al. 2017), a facilitators of interactions/mediators (Barakova et al. 2015; Gillet, van den Bos, and Leite 2020; Tozadore et al. 2022), or a companion (Song et al. 2021).

There is quite some supporting evidence for the role of a peer. Vogt et al. (2019) found that primary school children learned and retained second language vocabulary as effectively from a robot tutor framed as a co-learning peer as they did from a tablet application. Balkibekov et al. (2016) utilised the robot as a peer who is also learning a foreign language, demonstrating that primary school children could expand their vocabulary by playing a game against the robot. Framing the robot as a peer tutor has several advantages: it is perceived as more engaging and fun (Kanda et al. 2004), facilitates learning-by-teaching (Tanaka and Matsuzoe 2012), and strengthens the sense of relatedness between the learner and the robot. This, in turn, could aid children's intrinsic motivation (van Minkelen et al. 2020). However, when framing a robot as a peer, it is important that interactions are designed based on pedagogically well-established strategies to scaffold language learning (Vogt et al. 2017).

Davison et al. (2020) examined the development of primary school children's self-regulated learning over a four-month study, during which a social robot served as both tutor and instructor for various learning tasks. Belpaeme, Kennedy, et al. (2018) established guidelines for designing social robots as peer tutors for second language learning based on pedagogical and psycholinguistic theories. Hood, Lemaignan, and Dillenbourg (2015) applied a learning-by-teaching approach, having primary school children teach a robot to write, introducing a novel educational role for robots. Similarly, Sandygulova et al. (2020) used a robot to aid learners in transitioning from the Cyrillic to the Latin alphabet by teaching the robot to write. In the same way, Gargot et al. (2021) employed a similar learning-by-teaching approach with robots to motivate children with severe dysgraphia in handwriting training activities. Pareto, Ekström, and Serholt (2022) used a robot tutee as an inquisitive game companion for children in the context of an arithmetic game. Barakova et al. (2015) employed a social robot for content creation and co-design in therapy for children with autism spectrum disorder in an educational context. Neerincx et al. (2019) developed a robotic partner with which children performed educational activities to build step-by-step the competencies, relationships and autonomy for their diabetes self-management. Charisi et al. (2021) researched the role of a social robot in small group social interactions in a problem-solving setting with children five to seven years old. They found that specific robot behaviours can elicit more child–child task-related verbal interaction, leading to the improvement of metacognitive skills. Nasir, Bruno, and Dillenbourg (2024) investigated how a context-aware skill ignorant robot peer that tracks productive engagement state of learners, via multimodal behaviours, can facilitate learning in an open-ended collaborative algorithmic reasoning task. Norman et al. (2022) explored the use of a disagreeing collaborative robot to build shared solutions to advance the computational thinking skills of learners while Ligthart et al. (2023) investigated if a social robot that personalises conversation and scaffolds instructions leads to better learning outcomes when practicing math over time.

Looking at the evolution towards LLM-powered robots, the following research question remains a central concern for the field:

> (RQ3) Which social roles are most appropriate for introducing social robots in education that would support widespread use in the near and mid-term future?

3. Method

To answer the above research questions, we conducted a Delphi study with a panel of 16 experts. The Delphi method is a method for the 'systematic solicitation and aggregation of informed judgements from a group of experts on specific questions or issues' (Strauss and Zeigler 1975, 253). This technique has been successfully applied, among others, in the field of education (Hoang, Nhat, and Thi 2024), economy (Caldarelli 2024), engineering and technology (Bokrantz et al. 2017), health care (Denecke, May, and Rivera Romero 2024), and in robotics and machine learning (Engel and Dahlhaus 2021). There are several variations of the method, all of which emphasise on the expertise of participants and on orchestrating the survey of their opinion in such a way as to ensure mutual anonymity, to prevent confrontation and debate, but also iterations to allow experts to revise and refine their views.

While the Delphi method traditionally aims for a consensus between the experts there has been a considerable amount of criticism of how that consensus is reached. Ehringfeld et al. (2023) compared the majority, approval, Borda and range voting schemes for the specific use in Delphi studies. They concluded that despite that the most commonly used voting scheme is majority voting (i.e. when participants select their most preferred option, and the winner option (i.e. 'consensus') is the one that received the majority of the votes), it brushes over participants' deeper preference structures, i.e. no understanding is acquired of participants preferences besides their most preferred option (3). To overcome this issue, Ehringfeld et al. (2023) recommend using the Borda coding scheme as 'an appropriate voting scheme for a ranking-type Delphi study' (4). In this approach, participants rank all available options according to a given criterion, and points are assigned to each option based on the inverted values of the given ranks (i.e. the number of answer options minus a given rank). This way the ranking reflects the experts' collective and aggregated opinion, that not only shows the most preferred option, but gives further insights into deeper preference structure. The Borda count method is most frequently used in medical contexts (e.g. Crowther et al. 2025; Deniz and Orhan 2022; Fuster-RuizdeApodaca et al. 2025; Ou et al. 2024; Pandya-Wood 2022), but also in other fields such as decision sciences (e.g. Oufella 2024), economics (e.g. Ecer, Böyükaslan, and Hashemkhani Zolfani 2022) and technology (e.g. Shamsi, Zakerijenad, and Zaraifard 2025). Our study procedures were designed to preserve the aforementioned aims of a Delphi-type study while keeping a pragmatic workload that would encourage the participation of the experts in the study (Brady 2015).

The study has been approved by the Ethical Review Board of [University] on 20 November 2023 with approval nr ERB2023ID581.

3.1. Participants

The study was organised by four of the co-authors, who were all based on the [University] and working in social robotics and learning. The Delphi method (Okoli and Pawlowski 2004) is best conducted with 10 to 18 participants. We invited 17 experts, all of whom agreed to participate, however, the final sample consists data from 16 of them due to non-response. We invited established researchers within the field of social robotics and education; the inclusion criterion was that they should have been actively publishing in this field in the last five years. Experts were all in the scientific network of the four researchers coordinating the study. The experts were individually approached via e-mail, requesting their participation in the study given their expertise within the field. Participation was voluntary, and anonymity was strictly maintained throughout the process. After the completion of the Delphi procedures and the content analysis by three of the co-authors, the participating experts were invited to review the collected data and co-author this manuscript, and suggest improvements, including those who did not express the wish to co-author, could comment and correct on drafts of the manuscript. This approach was followed to ensure that the manuscript represents accurately a consensus view, while preserving anonymity among participants, and avoiding an extractive approach where experts's input would be used without them having the opportunity to receive academic credit for them.

During the first round of data collection, participants provided information about their background. 14 participants identified themselves as 'social robotics expert' while two selected the 'other' category. 10 respondents indicated that they were 'active in research and knowledgeable on social robotics research', three considered themselves as 'one of the thought leaders in the field', two selected the option of 'being currently active in social robotics research, but not fully abreast of the state of the art', and one indicated that 'they have done some research, but they do not track developments in the field currently'. Our panel consists of experts from Sweden (1), Germany (1), Switzerland (1), Italy (1), Belgium (1), Greece (1), U.S.A. (1), Australia (2), France (2), and The Netherlands (5), which further strengthens the external validity of the study findings (Susskind, McKearnan, and Thomas-Larmer 1999).

3.2. Procedure

This study adhered to the recommendations of Beiderbeck et al. (2021), and was structured into four distinct phases as follows:

1. We identified the research objectives and determined the study format.
2. We identified the expert panel and ran the first round of data collection: we gathered experts' opinions on the topic of research.
3. We carried out the second round of data collection with the aim of achieving consensus among the participants.
4. We analysed the collected data, identified areas of consensus, interpreted potential scenarios, and derived conclusions.

Both data collection rounds were conducted online, via an online survey platform. The first round of data collection took place between November 2023 and January 2024, and the second round of data collection between February and April 2024.

3.2.1. First study round

Along with information on the participating experts' background, in the first iteration of data collection, we asked the panel twice the following three open-ended questions, referring to the time frame of the upcoming year (three questions) and within the next five years (three questions). These time horizons were selected to separate what is imminent and an expected evolution from their educated guess regarding future development. The focus of the questions were based on the insights of the study of Barakova et al. (2023).

1. Given recent developments in generative AI and large language models, what do you see as potential opportunities for imminent practical applications of social robotics in education with the current state of the art in the next year / within 5 years?
2. Given recent developments in generative AI and large language models, what obstacles do you see that are preventing the widespread use of social robots in education currently and in the upcoming year / within 5 years?
3. What roles do you think social robots should adopt in the educational context in the upcoming year / within 5 years that would facilitate the widespread use of them? Think about contexts and roles where help from a social robot would be appreciated.

For these questions, we received a wide variety of responses, on which we have conducted inductive content analysis (Mayring 2015), resulting in a set of structured response options that informed the second round of data collection.

3.2.2. Second study round

The second study round was built on the input of the first round, and it was conducted with the aim to reach a consensus on the key issues. The expert panel was presented with response options related to opportunities, challenges, and robot roles. Specifically, we asked the expert panel for each aspect (i.e. opportunities, challenges, and robot roles):

1. Rank the identified topics according to their importance (rank order, 1 – most important, 10 – least important)
2. Indicate the feasibility of the identified topics within one year (6-step Likert-scale; 1 – Not feasible/likely at all, 6 – Very much feasible/likely)
3. Indicate the feasibility of the identified topics within 5 years for the opportunities, challenges, and the robot roles (6-step Likert-scale; 1 – Not feasible/likely at all, 6 – Very much feasible/likely)

Thus, experts answered these nine questions, each followed by an open-ended question where they could provide additional comments on their rankings/ratings.

3.3. Data analysis

The analysis process began with the qualitative content analysis responses from the first round of data collection, which involved open-ended questions regarding potential opportunities afforded, challenges faced, and roles to be fulfilled by social robots. In the first step of the inductive content analysis (Mayring 2015) the data was coded by four coders, who then grouped these codes into themes which are listed in Tables 1–3. The four coders worked independently, creating their own codes for the data. The first author synthesised the different categories into a list of non-overlapping opportunities and agreements that would cover all those identified by the four coders.

We conducted quantitative analysis of the data collected in the second round, where experts were asked to rank and rate items based on their perceived relevance and feasibility. For the quantitative analysis, we utilised SPSS version 27.0 and R Studio 1.1.453 software.

Consensus was calculated based on the ranks, following the Borda coding method. Accordingly, participants ranked all available options according to a given criterion, and points were assigned to each option based on the inverted values of the given ranks (i.e. the number of answer options minus a given rank). For calculating agreement we used Fleiss' kappa. Fleiss' kappa can have a value between −1 to +1. A negative kappa (κ) means that agreement among raters is worse than what would be expected by random chance, with −1 representing complete disagreement (no overlap in ratings) and 0 indicating agreement equal to that of random chance. Values above 0 indicate progressively stronger agreement beyond chance, up to +1, which reflects perfect agreement (all raters made identical judgments) (Landis and Koch 1977). For the interpretation of the kappa values the amount of raters and response options should be considered critically. In our case we expected moderate values given the large number of raters (i.e. 15 experts; in comparison with the usually used 2–3 raters) and the amount of items to be ranked (8–10 options, a.k.a. items; in comparison with the usually used 3–5 response options).

For the rating, the sample mean and standard deviations were calculated. These values were calculated and added to the results to enrich the study insights.

Table 1. Content analysis results for the first round of data collection. Opportunities.

Opportunities	Example quotes
Creating ethical and legal regulations for the safe use of robots and AI in education	'Developing regulations on ethical standard and policies for using social robots in education'
Reducing related costs (e.g. purchase, maintenance, training etc.)	'Their cost, maintenance and operation is very distinct in relation to other kinds of technology' 'Hopefully it will be cheaper and easier to apply robots on a larger scale.'
Increasing the transparency, predictability, and controllability of AI and LLMs	'A responsible use of generative AI requires secure (privacy preserved) and transparent data processing and storage,' 'Hopefully 5 years from now (…) generative-AI content will be more transparent/controllable/predictable'
Supporting learners to be able to transfer learning from one topic to another	'The ability to transfer the personalization model in various educational contexts'
Enabling multimodal input and integration with other devices and tools	'I believe we will see radical changes mainly with the use of LLMs but also with large multimodal models.' 'I think what is interesting is that generative AI are becoming (or are already) multimodal, and that social robots are also inherently interacting with the world around them in a multimodal way. It will be able to integrate its inputs through vision, speech recognition, touch, etc. and be better able to understand the world around it and the people present in there.'
Improving the user experience (e.g. believability, multiple language support, natural interactions etc.)	'The novel AI methods allow for the generation of a more fluent and "believable" interaction experience, which in turn is important to creating a responsive and effective HRI.'
Creating new didactics and pedagogies for the use of social robots in education	'Defining and describing a clear integration model in Education (including not only scenarios of use (clear impact in teaching and learning) but also the practicalities: how many, what is the cost, what does it mean in terms of maintenance).'
Creating content that supports long-term learning activities	'Their greatest potential I think will be in content generation to facilitate long-term interaction.'

4. Results

4.1. First iteration of open data collection

Despite the panel being asked separately to assess opportunities and challenges within two different time frames of one year and five years respectively, during the content analysis, the same categories emerged for the two time horizons indicating that at this stage, experts did not hold beliefs that differ substantially between the short and the mid-term. The content analysis resulted in eight categories for the opportunities (Table 1), ten categories for the challenges (Table 2), and ten categories for the robot roles (Table 3). It is noted how all the opportunities may apply generally to the implementation of LLMs in education, though the combination with social robotics does set different requirements to address the robot embodiment. Robots combine multiple sensing modalities and actuation, but also lead to different interactive experiences, emotional responses and attachment than

Table 2. Content analysis results for the first round of data collection. Challenges.

Challenges	Example quotes
Overcoming technological limitations (e.g. hardware, software etc.)	'related to technical issues: computational cost, social signal processing, appropriateness of responses, … '
Addressing the lack of legal regulations and ethical standards for the use of SRs in education (e.g. data handling, reproducibility, biases, privacy)	'Legal privacy aspects will probably be the first obstacle as these companies are not transparent about how they handle data, and the legal space is difficult to navigate in this regard.'
Addressing risk factors to ensure the safety of learners (e.g. securing infrastructure, improving reliability, preventing data leakage etc.)	'lack of control and predictability over the outcome of a generative AI, both in its factual correctness and its social appropriateness'
Decreasing related costs (e.g. purchase, maintenance, reparations etc.)	'For the coming decade I can see the following obstacles: the price and logistic challenges of deploying social robots in education … '
Improving supporting practices (e.g. training of stakeholders, technical support etc.)	'education of stakeholders to the use of AI'
Addressing the lack of pedagogy, theoretical grounding, and teaching methodologies for the use of SRs in education	'theoretical lack of new didactics and pedagogies for the use of SR in education'
Addressing the lack of educational content (e.g. to sustain multiple years of curricula sessions)	'Another obstacle is that it takes time to develop a feasible educational application using a robot in an educational setting that will make use of these new LLMs.' 'The main obstacles are how to organize their use (train and educate the teachers) and embed educational social robots in the teaching curricula of schools.'
Overcoming negative attitudes and increasing technology acceptance, including mistrust in AI	'Getting teachers on board. They mostly already have little time available, for example to prepare their classes, so they will find it hard to find time to learn how the robot works and maybe design content for it.'
Improving accessibility and inclusiveness (e.g. to decrease digital divide)	'The other important problem is the digital divide and how social robots can help disadvantaged students to develop and overcome the obstacles that might emerge from the lack of digital skills from their parents.'
Improving robustness across various contexts, making robots more adaptive and connected	'Enhancing the resilience of robots in educational environments to strengthen their robustness'

Table 3. Content analysis results for the first round of data collection. Robot roles.

Robot Roles	Example quotes
Teachers use the robot as a teaching assistant that complements the teacher's work in the classroom (e.g. acts as a storyteller, role-player, as a peer learner, teachable buddy, facilitates group interactions etc.)	'I think it [Gen AI and LLMs] will make the role of teaching assistant more feasible mainly, given that expectations are focused on conversational (speech-based) interaction.' 'They should be deployed with a focus on core curriculum topics such as math and language learning, where they can act as "teaching assistants" of teachers (complementing mainly for capacity issues).' 'An additional tool for the teacher in the classroom activities'
Learners use the robot as a conversational interface that can answer their questions and provide easy access to information and knowledge (e.g. knowledge source, explanatory answers for questions etc.)	'I believe that the mainly factually correct responses that generative-AI-supported robots can provide will further promote their use as knowledge providers' 'Using robots as knowledge conveyors – helping learners answer questions on open (or restricted) domains through Q&A'
Learners use the robot as a personalised tutor for self-study (e.g. homework study buddy)	'Using robot for supporting self-study' 'Personalized learning assistants'
Learners use the robot for emotional support (e.g. someone to open your heart to, to promote self-disclosure and self-assessment of psychological states such as stress or anxiety etc.)	'the use of social robots as an extra-curricular resource to promote the students' self-disclosure and self-assessment of their psychological states (stress, anxiety, depression, bullying) as a way to encourage students in need to address the problem properly (e.g. motivate them to seek help by a qualified professional, speak up with the teachers/parents …)'
Learners use the robot for companionship and for bonding (e.g. as a buddy)	'The use of LLMs to create an open-ended interaction between the student and the robot *not related to the educational topic at hand*. This serves to create a social bond between the student and the robot, and has been shown to be conducive to learning.'
Teachers use the robot to increase learners' interest in a topic (e.g. to increase attitudes towards STEM subjects)	'Using robot for increasing children's interest in a topic, as a motivational agent' 'a way to increase children's interest in the topic, via the novelty effect'
Teachers use the robot for assisting disadvantaged learners (e.g. learners with special needs)	'how social robots can help disadvantaged students to develop and overcome the obstacles that might emerge from the lack of digital skills from their parents.' 'Supporting special needs population'
Teachers use the robot to stimulate critical thinking, discussions, and debates (e.g. emphasising understanding over memorisation by repetition)	'robots designed to stimulate discussions and debates and enhance critical thinking (shift in paradigm, emphasizing understanding over memorization by repetition)'
Robots take over the teachers' work as a whole (in contrast with complementing it)	'teacher'
Robot adopts multiple roles and accompanies students in formal and informal educational settings (e.g. teachable peer during school classes, buddy on the way home, and personalised tutor / study buddy in the afternoon at home).	'I envision a robot that is with you in both formal and informal settings. E.g. it is practicing/learning with you in class, then at home helps you with homework'

interaction through the screen. The only exception to this remark pertains to developing content for longer term learning activities, which are not specific to robots. Similarly, all challenges can be extended to the application of AI in education, with the exception perhaps of developing educational content – this may exist in abundance outside the field of social robotics and would need to be adapted for use with robots. We note also an overlap between what experts considered an opportunity, presumably for research, and the challenges. With slight differences in emphasis and phrasing, ethical and legal aspects, pedagogy, technology development and developing content were considered both as challenges and opportunities.

4.2. Second round of data collection: reaching consensus

The individual rankings provided by experts regarding the importance of different opportunities were synthesised by calculating a rank order as discussed in Section 3.3. For the likelihood of attaining these opportunities and the feasibility of dealing with challenges, sample means of the experts' ratings were computed.

4.2.1. Ranking of opportunities

The ranking of opportunities is shown in Table 4. The overall agreement on the ranking is substantial and significant ($k = 0.042$, $p = 0.001$). According to the expert panel, the two most important opportunities are to have the ethical and legal regulations settled and to develop suitable didactics and pedagogies for the use of social robots in education.

4.2.2. Expected likelihood of realising opportunities in the upcoming year

The mean ratings of the different objectives regarding feasibility within 1 and 5 years are shown in Table 4. We note that the experts found that five out of the eight options are unlikely to be realised within the upcoming year: reducing related costs, creating content that supports long-term learning, developing new didactics, establishing ethical and legal regulations, and enhancing transparency. Our panel was positive about social robots becoming able to help learners

Table 4. Summary of results related to realising opportunities.

Topic	Rank	Expected likelihood to be realised within 1 year	Expected likelihood to be realised within 5 years
Creating ethical and legal regulations for safe use of robots and AI in education,	1 $k = 0.064$*	$M = 2.80$ (SD = 1.57) median = 2 (min = 1, max = 5)	$M = 4.40$ (SD = 1.24) median = 4 (min = 3, max = 6)
Creating new didactics and pedagogies for the use of social robots in education	2 $k = -0.012$	$M = 2.73$ (SD = 0.80) median = 3 (min = 2, max = 4)	$M = 4.27$ (SD = 0.96) median = 4 (min = 3, max = 6)
Increasing the transparency, predictability, and controllability of AI and LLMs	3 $k = 0.031$	$M = 2.87$ (SD = 1.25) median = 2 (min = 1, max = 5)	$M = 4.27$ (SD = 1.28) median = 4 (min = 2, max = 6)
Improving the user experience (e.g. believability, multiple language support, natural interactions etc.),	4 $k = -0.012$	$M = 3.87$ (SD = 0.74) median = 4 (min = 3, max = 5)	$M = 4.93$, (SD = 0.80) median = 5 (min = 4, max = 6)
Enabling multimodal input and integration with other devices	5 $k = -0.012$	$M = 3.67$ (SD = 0.82) median = 4 (min = 2, max = 5)	$M = 4.93$ (SD = 0.80) median = 5, (min = 3, max = 6)
Creating content that supports long-term learning activities,	6 $k = 0.107$*	$M = 2.67$ (SD = 1.29) median = 3 (min = 1, max = 5)	$M = 4.40$ (SD = 1.12) median = 5 (min = 2, max = 6)
Reducing related costs (e.g. purchase, maintenance, training etc.)	7 $k = 0.031$	$M = 2.27$ (SD = 1.22) median = 2 (min = 1, max = 4)	$M = 4.13$ (SD = 1.19) median = 4 (min = 2, max = 6)
Supporting learners to be able to transfer learning from one topic to another	8 $k = 0.140$*	$M = 3.33$ (SD = 1.50) median = 3 (min = 1, max = 6)	$M = 4.07$ (SD = 1.03) median = 4 (min = 2, max = 6)

Notes: K denotes the Kappa value for agreement on the rank. * indicates $p < 0.1$. Values below 3 are considered unlikely, values above 3 are considered likely to be realised within the given timeframe. The top and bottom three topics of the rankings and ratings are indicated with colour code: Top 1, Top 2, Top 3, Bottom 1, Bottom 2, Bottom 3. Please note that colours in the Table will not be visible in a printed version.

transfer knowledge from one topic to another, enabling multimodal input, and realising an improved user experience within the upcoming year.

The ratings show a clear contrast between the long-term, more challenging objectives and the more achievable, near-term goals. Objectives unlikely to be realised within the next year highlight the need for sustained effort, collaboration, and a longer timeline for success. In contrast, short-term goals reflect current advancements in robot technology and its educational applications. While progress in user experience and interaction is promising, addressing long-term challenges remains essential. Overall, the findings stress the importance of balancing immediate improvements with efforts to overcome long-term obstacles (e.g. ethical and legal regulations) for the widespread adoption of social robots in education. Given that we have 16 ratings only, the standard deviations of the ratings were not too large, with the exception of putting in place ethics and legal regulations for safe use, and the transfer of learning across topics – both of which may result as the outcome of prolonged efforts.

4.2.3. Expected likelihood of realising opportunities within five years

Looking at the 5-year horizon, our panel appeared to be more optimistic about what was achievable, as all opportunities were rated higher than the midpoint of the scale (i.e. $M \geq 3$). Experts were confident that the user experience will be improved, that multimodal input will be possible, that content will be created that supports long-term learning, and that ethical and legal regulations will be developed within five years. They concur in that didactics and pedagogies will emerge within this timeframe, that the transparency and predictability of AI and LLMs will improve sufficiently, that related costs will be reduced and that supporting learners to be able to transfer learning from one topic to another is feasible within the 5-year horizon. Perhaps surprisingly, given how difficult it is to predict the future, the standard deviations of the ratings were lower for the five year horizon than for the next year, perhaps indicating that the optimism for realising the opportunities identified was shared among the experts.

Comparing the ranking of the opportunities with the feasibility of those it appears that what should be the most important objectives to achieve (creating ethical and legal regulations and creating new pedagogies and didactics) were both rated unlikely within the upcoming year, and only probably likely within five years. A summary of results related to realising objectives are presented in Table 4.

4.2.4. Ranking of challenges to overcome

The ranking of challenges is shown in Table 5. The overall agreement on the ranking is substantial and significant ($k = 0.022$, $p = 0.031$). We note that the panel of experts considered that the two most important challenges to overcome within the field of social robots in education are to address the lack of ethical and legal regulations and to deal with technological limitations.

4.2.5. Expected likelihood of dealing with challenges within the upcoming year

The mean ratings of the different challenges to deal within one and five years are shown in Table 5. We note that the experts found that three out of the ten challenges are unlikely to be realised within the upcoming year: addressing the lack of ethical standards and legal regulations and improving robustness across various contexts. Our panel was slightly more positive about the social robotics field improving accessibility and inclusiveness, ensuring user safety, and establishing pedagogy and theoretical grounding.

Comparing these results with the ranking of the challenges to overcome we see that on one hand what should be the most important issue to deal with (i.e. Addressing the lack of legal regulations and ethical standards) was rated the least feasible within the upcoming year. On the other hand, what was ranked as the least important challenge to overcome within the upcoming year (i.e. *Improving accessibility* and inclusiveness) was rated the most feasible one. The ratings were very aligned among experts for most challenges, apart from the lack of legal regulations and ethical standards and overcoming technical limitations in the next year.

4.2.6. Expected likelihood of dealing with challenges within five years

The expert panel was optimistic about the feasibility of addressing challenges within five years, our expert panel appeared to be quite optimistic, as they rated all items above the midpoint of the scale (i.e. $M \geq 3$) on average. The two most feasible challenges to address within five years were *Improving supporting practices* and *Improving robustness* across various contexts.

A discrepancy was found between the perceived importance and feasibility of addressing several challenges. For example, the experts considered addressing the lack of legal regulations and ethical standards as the most important challenge to overcome, yet rated its feasibility as least feasible, even within the five-year timeframe. Conversely, while the expert panel ranked improving accessibility and inclusiveness as least

Table 5. Summary of results related to dealing with challenges.

Topic	Rank	Expected likelihood to be dealt with within 1 year	Expected likelihood to be dealt with within 5 years
Addressing the lack of legal regulations and ethical standards for the use of social robots in education	1 $k = 0.016$	$M = 2.73$ ($SD = 1.34$) median = 3 (min = 1, max = 6)	$M = 4.27$ ($SD = 0.88$) median = 4 (min = 3, max = 6)
Overcoming technological limitations	2 $k = 0.005$	$M = 3.00$ ($SD = 1.41$) median = 3 (min = 1, max = 6)	$M = 4.33$ ($SD = 0.90$) median = 4 (min = 3, max = 6)
Addressing risk factors to ensure the safety of learners	3 $k = 0.005$	$M = 3.20$ ($SD = 0.86$) median = 3 (min = 2, max = 5)	$M = 4.27$ ($SD = 0.80$) median = 4 (min = 3, max = 5)
Addressing the lack of pedagogy, theoretical grounding, and teaching methodologies for the use of social robots in education	4 $k = -0.005$	$M = 3.20$ ($SD = 1.01$) median = 4 (min = 1, max = 4)	$M = 4.33$ ($SD = 0.90$) median = 4 (min = 3, max = 6)
Addressing the lack of educational content	5 $k = -0.005$	$M = 3.20$ ($SD = 1.21$) median = 3 (min = 1, max = 5)	$M = 4.27$ ($SD = 0.96$) median = 5 (min = 2, max = 5)
Decreasing related costs	6 $k = 0.037$	$M = 2.73$ ($SD = 0.70$) median = 3 (min = 2, max = 4)	$M = 4.33$ ($SD = 0.82$) median = 4 (min = 3, max = 6)
Improving supporting practices	7 $k = 0.037$	$M = 3.07$ ($SD = 1.03$) median = 3 (min = 1, max = 5)	$M = 4.47$ ($SD = 0.83$) median = 5 (min = 3, max = 6)
Improving robustness across various contexts, making robots more adaptive and connected	8 $k = 0.048$	$M = 2.73$ ($SD = 0.80$) median = 3 (min = 2, max = 4)	$M = 4.40$ ($SD = 1.06$) median = 5 (min = 2, max = 6)
Overcoming negative attitudes and increasing technology acceptance, including mistrust in AI	9 $k = 0.016$	$M = 3.13$ ($SD = 1.06$) median = 3 (min = 2, max = 6)	$M = 4.20$ ($SD = 0.94$) median = 4 (min = 2, max = 5)
Improving accessibility and inclusiveness	10 $k = 0.069$*	$M = 3.27$ ($SD = 1.28$) median = 3 (min = 1, max = 6)	$M = 4.07$ ($SD = 0.96$) median = 4 (min = 2, max = 6)

Notes: *K* denotes the Kappa value for agreement on the rank. * indicates $p < 0.1$. Values below 3 are considered unlikely, values above 3 are considered likely to be realised within the given timeframe. The top and bottom three topics of the rankings and rating are indicated with colour code: Top 1, Top 2, Top 3, Bottom 1, Bottom 2, Bottom 3. Please note that colours in the Table will not be visible in a printed version.

important, it was rated with the highest feasibility scores within the upcoming year. Again, the ratings of the likelihood of addressing challenges were better aligned across experts for the five-year horizon than for the next year. A summary of results related to challenges to overcome are presented in Table 5.

4.2.7. Ranking of robot roles

The ranking of robot roles is shown in Table 6. The overall agreement on the ranking is substantial and significant ($k = 0.135$, $p < 0.001$). We note that the panel of experts considered the two most important robot roles to be fulfilled within the field of social robots in education to use robots as teaching assistants and personalised tutors.

4.2.8. Expected likelihood of robot roles to be implemented in the upcoming year

The mean ratings of the different robot roles to be implemented within one and years are shown in Table 6. We note that the experts found that five out of the ten roles are unlikely to be realised within the upcoming year: robots taking over teachers' work completely, being able to adapt to different contexts, being used as an emotional support or companion, and stimulating critical thinking and debates. Our panel was positive about social robots becoming able to increase learners' interest, providing access to information and knowledge, and assisting disadvantaged learners within the upcoming year.

Comparing these findings to the ranking of the robot roles we conclude that the expert panel ranked robots taking over teachers' work both the least important, and least feasible to be implemented within the upcoming year. However, the role of teaching assistant, while ranked as the most important role to be implemented, was rated only moderately feasible within the upcoming year. Notably, the opinions of experts were also most spread on this point as indicated by the standard deviation.

4.2.9. Expected likelihood of robot roles to be implemented in five years

Regarding the five-year timeframe, the expert panel ratings suggest that social robots are most likely to be used as conversational interfaces and to increase learners' interest. The panel was borderline sceptical about robots adopting multiple roles and was clearly skeptical whether robots will take over teachers' work completely.

Table 6. Summary of results related to robot roles.

Topic	Rank	Expected likelihood to be implemented within 1 year	Expected likelihood to be implemented within 5 years
Teachers using the robot as teaching assistant that complements the teacher's work in the classroom	1 $k = 0.206$*	$M = 3.29$ (SD = 1.33) median = 4 (min = 1, max = 5)	$M = 4.29$ (SD = 0.99) median = 4 (min = 2, max = 6)
Learners use the robot as a personalised tutor for self-study	2 $k = 0.048$	$M = 3.07$ (SD = 0.92) median = 3 (min = 2, max = 5)	$M = 4.14$ (SD = 0.66) median = 4 (min = 3, max = 5)
Learners use the robot as a conversational interface that can answer their questions and provide easy access to information and knowledge	3 $k = 0.143$*	$M = 3.86$ (SD = 1.17) median = 4 (min = 2, max = 6)	$M = 4.79$ (SD = 0.80) median = 5 (min = 3, max = 6)
Teachers use the robot for assisting disadvantaged learners	4 $k = 0.079$	$M = 3.57$ (SD = 1.16) median = 3 (min = 2, max = 6)	$M = 4.43$ (SD = 0.76) median = 4 (min = 3, max = 6)
Teachers use the robot to increase learners' interest in a topic	5 $k = 0.026$	$M = 4.43$ (SD = 0.94) median = 4 (min = 3, max = 6)	$M = 4.79$ (SD = 1.19) median = 5 (min = 2, max = 6)
Teachers use the robot to stimulate critical thinking, discussions, and debates	6 $k = -0.005$	$M = 2.86$ (SD = 1.29) median = 3 (min = 1, max = 5)	$M = 3.71$ (SD = 0.99) median = 3 (min = 2, max = 5)
Learners use the robot for emotional support	7 $k = 0.079$	$M = 2.36$ (SD = 0.84) median = 2 (min = 1, max = 4)	$M = 3.71$ (SD = 0.73) median = 4 (min = 3, max = 5)
Learners use the robot for companionship and for bonding	8 $k = 0.122$*	$M = 2.64$ (SD = 1.08) median = 2 (min = 1, max = 5)	$M = 3.57$ (SD = 0.94) median = 3 (min = 2, max = 5)
Robot adopts multiple roles and accompanies learners in formal and non-formal educational setting	9 $k = 0.058$	$M = 2.00$ (SD = 0.78), median = 2 (min = 1, max = 3)	$M = 3.00$ (SD = 1.04) median = 3 (min = 1, max = 5)
Robots take over teachers' work as a whole (in contrast with complementing it)	10 $k = 0.598$*	$M = 1.36$ (SD = 0.84), median = 1 (min = 1, max = 4)	$M = 1.50$ (SD = 0.94) median = 1 (min = 1, max = 4)

Notes: K denotes the Kappa value for agreement on the rank. * indicates $p < 0.05$. Values below 3 are considered unlikely, and values above 3 are considered likely to be implemented within the given timeframe. The top and bottom three topics of the rankings and ratings are indicated with colour code: Top 1, Top 2, Top 3, Bottom 1, Bottom 2, Bottom 3. Please note that colours in the Table will not be visible in a printed version.

Despite the experts ranking the role of teaching assistant as the most important for social robots in education, they believe it is only moderately likely to be widely adopted within the upcoming five years. Instead, the experts anticipate that robots will more likely be used for evoking learners' interest and learners using the robots as a conversational interface, which they had previously ranked 5th and 3rd in importance respectively.

Analysis of the data on opportunities, challenges, and social robot roles shows that the most important issues are not always the easiest to address. From these three investigated topics (opportunities, challenges and robot roles), the ranking of the robot roles was the most closely aligned with the feasibility for implementation in practice. A summary of the feasible robot roles is shown in Table 6.

5. Discussion

This paper investigated the perspectives of a panel of 16 experts in HRI and educational technology regarding the future of social robots in education in light of the significant innovation potential of GenAI and LLMs. Delphi method was employed to understand how experts align in their views over the main opportunities, challenges, and promising robot roles within the field for the short-term (one year) and mid-term (five year) horizons.

Opportunities. The experts considered the two most important opportunities in the field of social robots in education (i) to develop ethical frameworks and legal regulations and (ii) to develop didactics and pedagogies for using social robots in education. The need for guidelines for applying robots in education (e.g. Belpaeme, Kennedy, et al., 2018), and didactic content creation (e.g. Barakova et al. 2015) has been raised before, but with the advancement of GenAI, the expert panel emphasised the urgency of going a step further to develop legal frameworks next to a suitable pedagogy. Introducing advanced technology is not sufficient, effective integration requires careful consideration of how robots can be best used to facilitate learning, but also ensure originality and integrity, tempering the technology push by a cautious and thoughtful approach, adhering to the stringent standards observed in education. The creation of suitable educational content and pedagogical strategy design with the use of GenAI-powered robotics on its own is quickly re-focusing on personalisation, with adaptive learning paths tailored to individual student needs, pacing, and styles. The curricula now should utilise LLMs for generated feedback, and multimedia materials. As a result, teachers could be evolving into facilitators of AI-enhanced educational content creation, requiring professional development in AI literacy, ethical considerations, and pedagogical integration (Hu et al. 2025).

Furthermore, the experts do not expect the introduction of GenAI powered robots to be imminent, however, they consider it feasible within five years. Current regulations for HRI (Mahler 2024) focus on safety regarding the environment and the physical behaviour of the robot, including aspects such as its predictability, the force it can exercise, etc. Regarding social behaviours, the notion of adequacy is stipulated, which refers to robots adhering to social norms. This generic conception of adequacy may need to be re-interpreted and specialised for an educational environment and for dealing with underaged users, but this has not yet been realised. Therefore, while current regulations provide a foundational understanding of robot's physical and social safety, they do not yet offer specific enough guidelines tailored to the unique context of education and involvement of children in interactions with GenAI-powered robots.

Regarding achievable aims, the expert panel considered that within the short-term horizon, social robots will be able to support learners in transferring knowledge from one session and topic to another, will enable multimodal interaction, and an improved user experience will be achieved. We note that a year has already elapsed since the collection of data and while developments in LLMs are highly paced, the challenges that the field should address remain. Furthermore, there is no evidence yet that they have been overcome or that the objectives set for the field have already been achieved. This illustrates how the distance between technological developments and their practical implementation may be underestimated. On the mid-term horizon, the experts were confident that in addition to the previous, content will be created that supports long-term learning, ethical frameworks and legal regulations will be created, new didactics and pedagogies will emerge, the transparency and predictability of GenAI and LLMs will improve sufficiently, and related costs will be reduced. Earlier research studies (Kennedy, Baxter, and Belpaeme 2015; Vogt et al. 2019; Woo et al. 2021) that considered possible ways for implementing social robots in education partially align with the steps the expert panel considered necessary to take within the mid-term horizon. This alignment could indicate that the concerns raised are applicable more broadly in the field of social robotics and the application of GenAI in education.

Challenges. As just discussed, the most important opportunities found in the current study were to create ethical and legal regulations for the safe use of robots and AI in education and to develop new pedagogies and didactics for their effective integration. Unsurprisingly, these priorities align with two of the most important challenges identified by the experts, next to overcoming technological limitations. The ranking of opportunities and challenges aligned in terms of top priority, such as creating ethical frameworks and legal regulations for safe use of robots and AI in education / addressing the lack of it, were both ranked as number one. However, the experts appeared quite skeptical as they found most challenges to be dealt with within the short-term horizon borderline feasible. Accordingly, they rated improving accessibility and inclusiveness, addressing lack of pedagogy and theoretical grounding, and addressing risk factors to ensure safety as the top three most feasible challenges to deal with. In the mid-term horizon, improving supporting practices and improving robustness across various contexts were the two most feasible challenges to deal with. Some of these points have been raised before, though not in reference to GenAI-powered robots (Westlund et al. 2018). The current study helps organise them in a comprehensive overview also indicating their priority.

Interestingly, importance and feasibility are not aligned. Our experts ranked addressing the lack of legal regulations and ethical standards as the most important challenge to overcome, however, when rating its feasibility, even within the five-year timeframe, this challenge was rated among the least feasible to deal with. On the other hand, improving accessibility and inclusiveness was ranked as least important by the expert panel, however, it was rated with the highest feasibility scores within the short-term horizon. This perspective, however, contrasts sharply with evidence surrounding the socioeconomic and technical realities of integrating GenAI into educational settings. While the importance of accessibility and inclusiveness is fundamentally about equipping individuals to shape and become tomorrow's society and the use of GenAI-powered social robots is frequently projected as having the power to democratise access to help and high-quality tutoring, achieving true accessibility and inclusiveness faces significant constraints, primarily related to cost, the digital divide, and the complexity of specialised needs. First, the claims suggesting LLMs will broaden accessibility to high-quality education are not supported by evidence yet. Only limited capability LLMs are available in free tiers, meaning that more powerful tools are often paid services. This strongly suggests that high-quality, LLM-enhanced education may be restricted to an elite who can afford the necessary infrastructure and access, thus further increase the existing socio-economic gaps and widening the digital divide. Furthermore, high costs and maintenance may limit access to GenAI-based social robots in underfunded or rural regions, reducing the potential for equitable access in special education settings. In addition, LLMs designated for special education must be trained with specific data and the needs will vary among different impairment groups (e.g. autism spectrum disorder, dyslexia, hearing impairment), meaning that distinct LLMs must be developed and tailored for each group and integrated with appropriate social robots. This requires substantial domain-specific knowledge and labelled datasets. However, access to large, diverse datasets for training LLMs in sensitive settings is severely constrained by data privacy and ethical concerns, limiting the personalisation capabilities of the resulting robots. In addition, LLM robots may introduce additional barriers for non-native English speakers (Rutatola, Stroeken, and Belpaeme 2025), and research has found that LLM detectors can be disadvantaging non-native English writers (Jiang et al. 2024).

Roles. Addressing the potential roles social robots could take in the educational setting, the expert panel found that the most important role for the robot was that teachers use them as teaching assistant that complements the teacher's work in the classroom, followed by students use the robot as a personalised tutor for self-study, and students use the robot as a conversational interface that can answer their questions and provide easy access to information and knowledge. These findings, while aligned with earlier research regarding possible robot roles (e.g. Kennedy et al. 2016; Song et al. 2021), are novel, as none of the earlier research studied the importance of robot roles to be implemented first. Notably, GenAI technology has drastically changed the possible roles of robots in education. Regarding creating and implementing robot roles, our experts found that both on the short- and mid-term horizon, teachers using the robot to increase learners' interest, students using the robot as a conversational interface, and teachers using the robot for assisting disadvantaged learners the most feasible. While earlier research has extensively studied possible robot roles that are reflected in our findings (e.g. tutor (Kennedy et al. 2016), peers (Balkibekov et al. 2016; Nasir, Bruno, and Dillenbourg 2024; Park et al. 2017), facilitators of interactions (Barakova et al. 2015), or companions (Song et al. 2021)), it has not yet addressed the relative feasibility and implementability of those. Most earlier research casts robots in the role of tutors or buddies, which our experts indicated as one of the most important roles to implement, but only as moderately likely to be fulfilled within the upcoming five years. These roles, while consistent with earlier research, gain new significance with the introduction of LLM-enhanced tutors and buddies, now offer dynamic, context-aware interactions that can adapt to individual learners' needs, foster curiosity, and provide real-time access to vast knowledge bases. Reflecting these developments, the most feasible near-term applications identified by experts include using robots to spark learner interest, support disadvantaged students, and serve as intelligent conversational agents, roles that are now more implementable and impactful due to the integration of generative AI technologies. New studies show that such robots significantly boost learner engagement through storytelling, and personalised dialogue (Li et al. 2025; Tozadore et al. 2025), while also supporting disadvantaged students by tailoring content and scaffolding learning experiences (Li et al. 2025). Moreover, hybrid human–AI workflows, where teachers guide and moderate robot interactions, are emerging as the most effective model, combining pedagogical expertise with scalable automation. This evolution marks a transition from robots as static tools to intelligent, collaborative agents that enrich both teaching and learning.

6. Limitations and future work

Typical limitations of the Delphi method include that the iterative nature of the process can lead to participant fatigue, potentially resulting in less engaged or thoughtful responses in later rounds. We mitigated this threat by running a pre-determined number of rounds rather than iterating until reaching consensus on the conclusions which could have helped reduce the level of disagreement. To compensate for this limitation, we opened up the manuscript to all participants aiming to address their comments and that they stand by the conclusions and recommendations made. Moreover, the method is heavily dependent on the expertise and judgment of the selected panel, which introduces the risk of bias if the panel lacks sufficient diversity in disciplinary expertise, geographic representation, or professional experience. This panel is largely European and overall, from relatively richer countries which may limit how these results translate to other regions. The implementation of costly robots and GenAI in under-resourced educational systems may be too remote an ambition, and while our experts recognised challenges regarding current digital divide, the perspectives of researchers studying the global south have not been captured in this study. The panel is very multi-disciplinary, though the shared interest in social robotics may also bring about enthusiasm and optimism for the field.

To enhance the robustness and generalisability of findings, future research could benefit from the inclusion of larger and more diverse expert panels including stakeholders such as teachers, school directors, and parents. Ensuring a broader range of perspectives, including considerations of geographic, cultural and gender diversity (Conti et al. 2019), would enable a more nuanced understanding of the challenges and opportunities associated with integrating robots into educational contexts and their relative advantages and limitations compared to other technological developments. Additionally, the conclusions from the Delphi study could be tested in pilot implementation studies or broader opinion surveys. For example, pilot studies examining the real-world implementation of robots in classrooms and a thorough meta-analysis on the impact of social robots in education could yield valuable evidence to corroborate or refine the forecast of the experts. By integrating expert consensus with empirical insights, future research could develop a more comprehensive and actionable framework for the effective deployment of social robots in education. To an extent, our experts were challenged to make predictions for the future, at least with respect to the feasibility of meeting opportunities and addressing challenges, and such predictions are by their nature uncertain. However, the purpose of this study has not been a future prediction as such but getting clarity as to the priorities of the field in order to provide input to researchers prioritising different research goals and to decision makers regarding their own roadmaps and plans for introducing social robots in education settings.

7. Conclusion

In conclusion, while LLM-powered social robots present exciting opportunities for enhancing educational experiences through personalised learning, collaboration, and emotional support, their widespread adoption remains limited by regulatory, technical, financial, ethical and practical limitations. Our study highlighted both the potential and the challenges of integrating robots into educational environments. Specifically, we found that while the creation of ethical and legal regulations and the development of new pedagogies were seen as the most important opportunities, they were rated unlikely to be achieved within the upcoming year and only probably likely within five years. Despite this high importance, the experts considered the least feasible challenges to deal with, even within the five-year timeframe. This discord between what is considered most important and what is deemed feasible suggests a need for focused research efforts in these areas. The second major finding is about the future roles of robots in education in the GenAI era: teachers using the robot to increase learners' interest, students using the robot as a conversational interface to knowledge sources, and teachers using the robot for assisting disadvantaged learners were found as the most feasible roles. To fully realise the benefits of social robots, future efforts must focus on overcoming these limitations through targeted research and improved collaboration among roboticists, educators, and policymakers. Addressing issues like ethical and legal regulations, suitable pedagogies, scalability, AI sophistication, teacher training, and privacy concerns will be crucial to making robots a valuable and accessible tool in education. With continued development and evaluation, social robots have the potential to enhance educational practices and foster more engaging, inclusive, and effective learning environments. There are many barriers towards realising this aim and new questions arise regarding how to integrate robots in education. While robots may eventually become suitable tools to support learning, the question remains if we need social robots to transform educational practices or simply to complement educational practices where appropriate.

Acknowledgements

We are thankful for our panel of experts for sharing their knowledge with us.

Author contributions

CRediT: **Gabriella Tisza:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing; **Panos Markopoulos:** Conceptualization, Formal analysis, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing; **Sofia Serholt:** Writing – original draft, Writing – review & editing; **Jauwairia Nasir:** Writing – original draft, Writing – review & editing; **Omar Mubin:** Writing – original draft, Writing – review & editing; **Adriana Tapus:** Writing – original draft, Writing – review & editing; **Salvatore Anzalone:** Writing – original draft, Writing – review & editing; **Koen V. Hindriks:** Writing – original draft, Writing – review & editing; **Paul A. Vogt:** Writing – original draft, Writing – review & editing; **Vasiliki Charisi:** Writing – original draft, Writing – review & editing; **Emiel Krahmer:** Writing – original draft, Writing – review & editing; **Mark A. Neerincx:** Writing – original draft, Writing – review & editing; **Kaisa Väänänen:** Conceptualization, Formal analysis, Validation; **Daniela Conti:** Writing – original draft, Writing – review & editing; **Jan de Wit:** Writing – review & editing; **Emilia I. Barakova:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

Data is available from the first author upon reasonable request given the potentially sensitive nature of the data.

ORCID

Kaisa Väänänen http://orcid.org/0000-0002-3565-6021
Emilia I. Barakova http://orcid.org/0000-0001-5688-4878

References

Addlesee, A., N. Cherakara, N. Nelson, D. Hernández García, N. Gunson, W. Sieińska, M. Romeo, C. Dondrup, and O. Lemon. 2024. “A Multi-party Conversational Social Robot Using LLMs.” In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 1273–1275, Boulder, CO. https://doi.org/10.1145/3610978.3641112.

Allgeuer, P., H. Ali, and S. Wermter. 2024. “When Robots Get Chatty: Grounding Multimodal Human-Robot Conversation and Collaboration.” In *Artificial Neural Networks and Machine Learning – ICANN 2024. ICANN 2024. Lecture Notes in Computer Science*, Vol 15019, edited by M. Wand, K. Malinovská, J. Schmidhuber, and I. V. Tetko, 306–321. Cham: Springer. https://doi.org/10.1007/978-3-031-72341-4_21.

Alnajjar, F., C. Bartneck, P. Baxter, T. Belpaeme, M. L. Cappuccio, C. Di Dio, F. Eyssel, et al. 2021. *Robots in Education: An Introduction to High-Tech Social Agents, Intelligent Tutors, and Curricular Tools*. Routledge. https://doi.org/10.4324/9781003142706.

Alnajjar, F., M. L. Cappuccio, O. Mubin, R. Arshad, and S. Shahid. 2020. “Humanoid Robots and Autistic Children: A Review on Technological Tools to Assess Social Attention and Engagement.” *International Journal of Humanoid Robotics* 17 (6): 2030001. https://doi.org/10.1142/S0219843620300019.

Atuhurra, J. 2024. “Leveraging Large Language Models in Human-Robot Interaction: A Critical Analysis of Potential and Pitfalls.” https://doi.org/10.48550/arXiv.2405.00693.

Balkibekov, K., S. Meiirbekov, N. Tazhigaliyeva, and A. Sandygulova. 2016. “Should Robots Win or Lose? Robot’s Losing Playing Strategy Positively Affects Child Learning.” In *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016*, 706–711, New York, USA. https://doi.org/10.1109/ROMAN.2016.7745196.

Barakova, E., K. Väänänen, K. Kaipainen, and P. Markopoulos. 2023. “Benefits, Challenges and Research Recommendations for Social Robots in Education and Learning: A Meta-review.” In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2555–2561, Busan, Korea. https://doi.org/10.1109/RO-MAN57019.2023.10309345.

Barakova, E. I., P. Bajracharya, M. Willemsen, T. Lourens, and B. Huskens. 2015. “Long-term LEGO Therapy with Humanoid Robot for Children with ASD.” *Expert Systems* 32 (6): 698–709. https://doi.org/10.1111/exsy.12098.

Beiderbeck, D., N. Frevel, H. A. von der Gracht, S. L. Schmidt, and V. M. Schweitzer. 2021. "Preparing, Conducting, and Analyzing Delphi Surveys: Cross-Disciplinary Practices, New Directions, and Advancements." *MethodsX* 8:101401. https://doi.org/10.1016/j.mex.2021.101401.

Belpaeme, T., J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka. 2018. "Social Robots for Education: A Review." *Science Robotics* 3 (21): 1–10. https://doi.org/10.1126/scirobotics.aat5954.

Billing, E., J. Rosén, and M. Lamb. 2023. "Language Models for Human-Robot Interaction." In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 905–906, Stockholm, Sweden. https://doi.org/10.1145/3568294.3580040.

Bokrantz, J., A. Skoogh, C. Berlin, and J. Stahre. 2017. "Maintenance in Digitalised Manufacturing: Delphi-Based Scenarios for 2030." *International Journal of Production Economics* 191:154–169. https://doi.org/10.1016/j.ijpe.2017.06.010.

Brady, S. R. 2015. "Utilizing and Adapting the Delphi Method for Use in Qualitative Research." *International Journal of Qualitative Methods* 14 (5): 1–6. https://doi.org/10.1177/1609406915621381.

Bravo Perucho, A., and M. Alimardani. 2023. "Social Robots in Secondary Education." In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 355–359, Stockholm, Sweden. https://doi.org/10.1145/3568294.3580105.

Brown, L., R. Kerwin, and A. M. Howard. 2013. "Applying Behavioral Strategies for Student Engagement Using a Robotic Educational Agent." In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 4360–4365. Manchester, UK. https://doi.org/10.1109/SMC.2013.744.

Caines, A., L. Benedetto, S. Taslimipoor, C. Davis, Y. Gao, O. Andersen, Z. Yuan, et al. 2023. "On the Application of Large Language Models for Language Teaching and Assessment Technology." https://doi.org/10.48550/arXiv.2307.08393.

Caldarelli, G. 2024. "Expert Perspectives on Blockchain in the Circular Economy: A Delphi Study with Industry Specialists." *Journal of Cleaner Production* 465:142781. https://doi.org/10.1016/j.jclepro.2024.142781.

Charisi, V., L. Merino, M. Escobar, F. Caballero, R. Gomez, and E. Gomez. 2021. "The Effects of Robot Cognitive Reliability and Social Positioning on Child-Robot Team Dynamics." In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 9439–9445. Xi'an, China. https://doi.org/10.1109/ICRA48506.2021.9560760.

Charpentier, S., M. Chetouani, I. Truck, D. Cohen, and S. M. Anzalone. 2022. "Social Robots in Learning Scenarios: Useful Tools to Improve Students' Attention or Potential Sources of Distraction?" In *Social Robotics. ICSR 2022*, 124–134. Cham: Springer. https://doi.org/10.1007/978-3-031-24670-8_12.

Chou, H. S., L. T. Thong, H. S. J. Chew, and Y. Lau. 2023. "Barriers and Facilitators of Robot-Assisted Education in Higher Education: A Systematic Mixed-Studies Review." *Technology, Knowledge and Learning* 28 (2): 477–516. https://doi.org/10.1007/s10758-022-09637-3.

Conti, D., A. Cattani, S. Di Nuovo, and A. Di Nuovo. 2019. "Are Future Psychologists Willing to Accept and Use a Humanoid Robot in Their Practice? Italian and English Students' Perspective." *Frontiers in Psychology* 10:1–13. https://doi.org/10.3389/fpsyg.2019.02138.

Crowther, G., R. Dunning, G. Russell, E. Wolverson, and B. R. Underwood. 2025. "Dementia In-patient Units in Psychiatric Hospitals: Research Priority Setting." *BJPsych Bulletin* 49 (4): 228–234. https://doi.org/10.1192/bjb.2024.42.

Dai, W., J. Lin, F. Jin, T. Li, Y.-S. Tsai, D. Gasevic, and G. Chen. 2023. "Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT." https://doi.org/10.35542/osf.io/hcgzj.

Davison, D. P., F. M. Wijnen, V. Charisi, J. van der Meij, V. Evers, and D. Reidsma. 2020. "Working with a Social Robot in School." In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 63–72. Cambridge, UK. https://doi.org/10.1145/3319502.3374803.

Denecke, K., R. May, and O. Rivera Romero. 2024. "Potential of Large Language Models in Health Care: Delphi Study." *Journal of Medical Internet Research* 26:e52399. https://doi.org/10.2196/52399.

Deng, C., S. Lai, C. Zhou, M. Bao, J. Yan, H. Li, L. Yao, and Y. Wang. 2024. "ASD-Chat: An Innovative Dialogue Intervention System for Children with Autism Based on LLM and VB-MAPP." https://doi.org/10.48550/arXiv.2409.01867.

Denis, B., and S. Hubert. 2001. "Collaborative Learning in an Educational Robotics Environment." *Computers in Human Behavior* 17 (5-6): 465–480. https://doi.org/10.1016/S0747-5632(01)00018-8.

Deniz, N., and E. O. Orhan. 2022. "Development of a Multi-criteria Decision-Making-Based Assessment Model for Dental Material Selection: Engine-Driven Nickel-Titanium Instruments Case Study." *Clin Oral Invest* 26:2645–2659. https://doi.org/10.1007/s00784-021-04234-7.

Ecer, F., A. Böyükaslan, and S. Hashemkhani Zolfani. 2022. "Evaluation of Cryptocurrencies for Investment Decisions in the Era of Industry 4.0: A Borda Count-Based Intuitionistic Fuzzy Set Extensions EDAS-MAIRCA-MARCOS Multi-criteria Methodology." *Axioms* 11 (8): 404. https://doi.org/10.3390/axioms11080404.

Ehringfeld, I., P. Steinhoff, S. Störk, A. Whittal, and O. R. Herber. 2023. "The Voting Scheme Determines the Winner: Considering Possible Options for Designing a Delphi Study." *International Journal of Qualitative Methods* 22:1–8. https://doi.org/10.1177/16094069221148871.

Encarnação, P., T. Leite, C. Nunes, M. Nunes da Ponte, K. Adams, A. Cook, A. Caiado, J. Pereira, G. Piedade, and Ribeiro, M. 2017. "Using Assistive Robots to Promote Inclusive Education." *Disability and Rehabilitation: Assistive Technology* 12 (4): 352–372. https://doi.org/10.3109/17483107.2016.1167970.

Engel, U., and L. Dahlhaus. 2021. "Data Quality and Privacy Concerns in Digital Trace Data." In *Handbook of Computational Social Science,* Vol. 1, edited by U. Engel, A. Quan-Haase, S. X. Liu, and L. Lyberg 343–362. Routledge. London, UK. https://doi.org/10.4324/9781003024583-23.

Fuster-RuizdeApodaca, M. J., T. M. White, C. Prats-Silvestre, F. P. Holgado-Tello, and J. W. Lazarus. 2025. "A Metric to

Measure Long-Term Well-Being of People Living with HIV in Spain." *BMC Medicine* 23:460. https://doi.org/10.1186/s12916-025-04300-x.

Gargot, T., T. Asselborn, I. Zammouri, J. Brunelle, W. Johal, P. Dillenbourg, D. Archambault, M. Chetouani, D. Cohen, and S. M. Anzalone. 2021. "'It Is Not the Robot Who Learns, It Is Me.' Treating Severe Dysgraphia Using Child–Robot Interaction." *Frontiers in Psychiatry* 12:1–11. https://doi.org/10.3389/fpsyt.2021.596055.

Gentile, M., G. Città, S. Perna, and M. Allegra. 2023. "Do We Still Need Teachers? Navigating the Paradigm Shift of the Teacher's Role in the AI Era." *Frontiers in Education* 8:1–14. https://doi.org/10.3389/feduc.2023.1161777.

Gillet, S., W. van den Bos, and I. Leite. 2020. "A Social Robot Mediator to Foster Collaboration and Inclusion among Children." In *Robotics: Science and Systems XVI*. Corvalis, Oregon, USA. https://doi.org/10.15607/RSS.2020.XVI.103.

Hei, X., H. Zhang, and A. Tapus. 2023. "Robots in Education: Influence of Regulatory Focus Theory." In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2562–2568. Busan, South Korea. https://doi.org/10.1109/RO-MAN57019.2023.10309599.

Hei, X., H. Zhang, and A. Tapus. 2024. "Exploring Help-Seeking Behavior, Performance, and Cognitive Load in Individual Tutoring: A Comparative Study between Human Tutors and Social Robots." In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, 1035–1041. Pasadena, CA, USA. https://doi.org/10.1109/RO-MAN60168.2024.10731328.

Hoang, Cam-Tu Tran, Phuong Vu Nhat, and Huong-Thao Doan Thi. 2024. "Using the Delphi Methodology to Develop Technology Criteria to Assess e-Learning Readiness in Higher Education Institutions." *Journal of Science and Technology* 7 (1): 114–125. https://doi.org/10.55401/p0a42k67.

Höhn, S., J. Nasir, A. Paikan, P. Ziafati, and E. André. 2024. "Using Large Language Models for Robot-Assisted Therapeutic Role-Play: Factuality Is Not Enough!." In *ACM Conversational User Interfaces 2024*, 1–6. Luxembourg city, Luxembourg. https://doi.org/10.1145/3640794.3665886.

Höhn, S., J. Nasir, D. C. Tozadore, A. Paikan, P. Ziafati, and E. André. 2024. "Beyond Pretend-Reality Dualism: Frame Analysis of LLM-Powered Role Play with Social Agents." In *Proceedings of the 12th International Conference on Human-Agent Interaction*, 393–395. Tsukuba, Japan. https://doi.org/10.1145/3687272.3690894.

Honig, S., and T. Oron-Gilad. 2018. "Understanding and Resolving Failures in Human-Robot Interaction: Literature Review and Model Development." *Frontiers in Psychology* 9:1–21. https://doi.org/10.3389/fpsyg.2018.00861.

Hood, D., S. Lemaignan, and P. Dillenbourg. 2015. "When Children Teach a Robot to Write: An Autonomous Teachable Humanoid Which Uses Simulated Handwriting." In *ACM/IEEE International Conference on Human-Robot Interaction*, 83–90. Portland, USA. https://doi.org/10.1145/2696454.2696479.

Hu, B., J. Zhu, Y. Pei, and X. Gu. 2025. "Exploring the Potential of LLM to Enhance Teaching Plans through Teaching Simulation." *Npj Science of Learning* 10 (1): 7. https://doi.org/10.1038/s41539-025-00300-x.

Huskens, B., R. Verschuur, J. Gillesen, R. Didden, and E. 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): 345–356. https://doi.org/10.3109/17518423.2012.739212.

Irpan, A., A. Herzog, A. T. Toshev, A. Zeng, A. Brohan, B. A. Ichter, B. David, et al. 2022. "Do as I Can, Not as I Say: Grounding Language in Robotic Affordances." arXiv preprint arXiv:2204.01691.

Janssens, R., E. Verhelst, G. A. Abbo, Q. Ren, M. J. P. Bernal, and T. Belpaeme. 2024. "Child Speech Recognition in Human-Robot Interaction: Problem Solved?" https://doi.org/10.48550/arXiv.2404.17394.

Jiang, Y., J. Hao, M. Fauss, and C. Li. 2024. "Detecting ChatGPT-Generated Essays in a Large-Scale Writing Assessment: Is There a Bias against Non-native English Speakers?" *Computers & Education* 217:105070. https://doi.org/10.1016/j.compedu.2024.105070.

Kanda, T., T. Hirano, D. Eaton, and H. Ishiguro. 2004. "Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial." *Human–Computer Interaction* 19 (1–2): 61–84. https://doi.org/10.1207/s15327051hci1901&2_4.

Kanda, T., M. Shimada, and S. Koizumi. 2012. "Children Learning with a Social Robot." In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, 351–358. Boston, MA, USA. https://doi.org/10.1145/2157689.2157809.

Kasneci, E., K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, et al. 2023. "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education." *Learning and Individual Differences* 103:102274. https://doi.org/10.1016/j.lindif.2023.102274.

Kennedy, J., P. Baxter, and T. Belpaeme. 2015. "The Robot Who Tried Too Hard." In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 67–74. Portland, USA. https://doi.org/10.1145/2696454.2696457.

Kennedy, J., P. Baxter, E. Senft, and T. Belpaeme. 2016. "Social Robot Tutoring for Child Second Language Learning." In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 231–238. Christchurch, NZ. https://doi.org/10.1109/HRI.2016.7451757.

Kim, C. Y., C. P. Lee, and B. Mutlu. 2024. "Understanding Large-Language Model (LLM)-Powered Human-Robot Interaction." In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 371–380. Boulder, CO, USA. https://doi.org/10.1145/3610977.3634966.

Konijn, E. A., B. Jansen, V. Mondaca Bustos, V. L. N. F. Hobbelink, and D. Preciado Vanegas. 2022. "Social Robots for (Second) Language Learning in (Migrant) Primary School Children." *International Journal of Social Robotics* 14 (3): 827–843. https://doi.org/10.1007/s12369-021-00824-3.

Landis, J. R., and G. G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33 (1): 159–174. https://doi.org/10.2307/2529310.

Lemaignan S., N. Newbutt, L. Rice, J. Daly, and V. Charisi V. 2021. "UNICEF Guidance on AI for Children: Application to the Design of a Social Robot for and with Autistic Children." In: *Proceedings of the IROS2021 Workshop: The Roles of Robotics in Achieving the UN's Social Development Goals*. Prague, Chech Republic.

Lemaignan, S., N. Newbutt, L. Rice, and J. Daly. 2024. "'It's Important to Think of Pepper as a Teaching Aid or Resource External to the Classroom': A Social Robot in a School for Autistic Children." *International Journal of Social Robotics* 16 (6): 1083–1104. https://doi.org/10.1007/s12369-022-00928-4.

Li, J., F. Schijve, S. Li, Y. Yang, J. Hu, and E. Barakova. 2025. "Towards Emotion Co-regulation with LLM-Powered Socially Assistive Robots: Integrating LLM Prompts and Robotic Behaviors to Support Parent-Neurodivergent Child Dyads." In *Proceedings of 025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2025)*, 9223–9230, Hangzhou, China, https:doi.org/10.1109/IROS60139.2025.11246217.

Ligthart, M. E. U., S. M. de Droog, M. Bossema, L. Elloumi, K. Hoogland, M. H. J. Smakman, K. V. Hindriks, and S. Ben Allouch. 2023. "Design Specifications for a Social Robot Math Tutor." In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 321–330. https://doi.org/10.1145/3568162.3576957.

Mahadevan, K., J. Chien, N. Brown, Z. Xu, C. Parada, F. Xia, A. Zeng, L. Takayama, and D. Sadigh. 2024. "Generative Expressive Robot Behaviors Using Large Language Models." In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 482–491. Boulder, CO, USA. https://doi.org/10.1145/3610977.3634999.

Mahler, T. 2024. "Smart Robotics in the EU Legal Framework: The Role of the Machinery Regulation." *Oslo Law Review* 11 (1): 1–18. https://doi.org/10.18261/olr.11.1.5.

Mann, J. A., B. A. MacDonald, I.-H. Kuo, X. Li, and E. Broadbent. 2015. "People Respond Better to Robots than Computer Tablets Delivering Healthcare Instructions." *Computers in Human Behavior* 43:112–117. https://doi.org/10.1016/j.chb.2014.10.029.

Markelius, A. 2024. "An Empirical Design Justice Approach to Identifying Ethical Considerations in the Intersection of Large Language Models and Social Robotics." https://doi.org/10.48550/arXiv.2406.06400.

Mayring, P. 2015. "Qualitative Content Analysis: Theoretical Background and Procedures." In *Approaches to Qualitative Research in Mathematics Education: Advances in Mathematics Education*, edited by A. Bikner-Ahsbahs, C. Knipping, and N. Presmeg, 365–380. Dordrecht: Springer. https://doi.org/10.1007/978-94-017-9181-6_13.

Michaelis, J. E., and B. Mutlu. 2018. "Reading Socially: Transforming the In-home Reading Experience with a Learning-Companion Robot." *Science Robotics* 3 (21): 1–11. https://doi.org/10.1126/scirobotics.aat5999.

Mubin, O., C. J. Stevens, S. Shahid, A. A. Mahmud, and J.-J. Dong. 2013. "A Review of the Applicability of Robots in Education." *Technology for Education and Learning* 1 (1): 1–9. https://doi.org/10.2316/Journal.209.2013.1.209-0015.

Nasir, J., M. Abderrahim, B. Bruno, and P. Dillenbourg. 2023. "An HMM-Based Real-Time Intervention Methodology for a Social Robot Supporting Learning." In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2204–2211. Busan, South Korea. https://doi.org/10.1109/RO-MAN57019.2023.10309430.

Nasir, J., B. Bruno, and P. Dillenbourg. 2024. "Social Robots as Skilled Ignorant Peers for Supporting Learning." *Frontiers in Robotics and AI* 11:1–19. https://doi.org/10.3389/frobt.2024.1385780.

Nasir, J., U. Norman, B. Bruno, and P. Dillenbourg. 2020. "When Positive Perception of the Robot Has No Effect on Learning." In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 313–320. Naples, Italy. https://doi.org/10.1109/RO-MAN47096.2020.9223343.

Neerincx, M. A., W. van Vught, O. Blanson Henkemans, E. Oleari, J. Broekens, R. Peters, F. Kaptein, et al. 2019. "Socio-Cognitive Engineering of a Robotic Partner for Child's Diabetes Self-management." *Frontiers in Robotics and AI* 6:1–13. https://doi.org/10.3389/frobt.2019.00118.

Nichols, E., L. Gao, and R. Gomez. 2020. "Collaborative Storytelling with Large-Scale Neural Language Models." In *Motion, Interaction and Games*, 1–10. Virtual Event, SC, USA. https://doi.org/10.1145/3424636.3426903.

Norman, U., A. Chin, B. Bruno, and P. Dillenbourg. 2022. "Efficacy of a 'Misconceiving' Robot to Improve Computational Thinking in a Collaborative Problem Solving Activity: A Pilot Study." In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1413–1420. Naples, Italy. https://doi.org/10.1109/RO-MAN53752.2022.9900775.

Obaid, M., R. Aylett, W. Barendregt, C. Basedow, L. J. Corrigan, L. Hall, A. Jones, et al. 2018. "Endowing a Robotic Tutor with Empathic Qualities: Design and Pilot Evaluation." *International Journal of Humanoid Robotics* 15 (6): 1850025. https://doi.org/10.1142/S0219843618500251.

Okoli, C., and S. D. Pawlowski. 2004. "The Delphi Method as a Research Tool: An Example, Design Considerations and Applications." *Information & Management* 42 (1): 15–29. https://doi.org/10.1016/j.im.2003.11.002.

Ou, C., Z. Daly, M. Carter, W. A. Hall, E. Z. Zusman, A. Russolillo, S. Duffy, and E. Jenkins. 2024. "Developing Consensus to Enhance Perinatal Mental Health through a Model of Integrated Care: Delphi Study." *PLoS One* 19 (5): e0303012. https://doi.org/10.1371/journal.pone.0303012.

Oufella, S. 2024. "Hybrid Use of Borda Count and PROMETHEE Method for Maintenance Strategy Selection Problem." *Foundations of Computing and Decision Sciences* 49 (2): 139–160. https://doi.org/10.2478/fcds-2024-0009.

Pandya-Wood, R. 2022. "Sequential Research to Evaluate the Impact of Patient and Public Involvement on Cancer Research Outcomes: Using Interviews, Stimulus Material and a Modified Delphi Technique." *International Journal of Qualitative Methods* 21:1–15. https://doi.org/10.1177/16094069221081606.

Pareto, L., S. Ekström, and S. Serholt. 2022. "Children's Learning-by-Teaching with a Social Robot versus a Younger Child: Comparing Interactions and Tutoring Styles." *Frontiers in Robotics and AI* 9:1–17. https://doi.org/10.3389/frobt.2022.875704.

Park, E., K. J. Kim, and A. P. del Pobil. 2011. "The Effects of a Robot Instructor's Positive vs. Negative Feedbacks on Attraction and Acceptance towards the Robot in Classroom." In *Social Robotics. ICSR 2011*, 135–141. Berlin: Springer. https://doi.org/10.1007/978-3-642-25504-5_14.

Park, H. W., R. Rosenberg-Kima, M. Rosenberg, G. Gordon, and C. Breazeal. 2017. "Growing Growth Mindset with a Social Robot Peer." In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 137–145. Vienna, Austria. https://doi.org/10.1145/2909824.3020213.

Patel, T., and O. Scharenborg. 2024. "Improving End-to-End Models for Children's Speech Recognition." *Applied Sciences* 14 (6): 2353. https://doi.org/10.3390/app14062353.

Prather, J., P. Denny, J. Leinonen, B. A. Becker, I. Albluwi, M. Craig, H. Keuning, et al. 2023. "The Robots Are Here: Navigating the Generative AI Revolution in Computing Education." In *Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education*, 108–159. Turku, Finland. https://doi.org/10.1145/3623762.3633499.

Ramachandran, A., C.-M. Huang, and B. Scassellati. 2017. "Give Me a Break!." In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 146–155. Vienna, Austria. https://doi.org/10.1145/2909824.3020209.

Ramachandran, A., S. S. Sebo, and B. Scassellati. 2019. "Personalized Robot Tutoring Using the Assistive Tutor POMDP (AT-POMDP)." *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (1): 8050–8057. https://doi.org/10.1609/aaai.v33i01.33018050.

Riedmann, A., P. Schaper, and B. Lugrin. 2024. "Integration of a Social Robot and Gamification in Adult Learning and Effects on Motivation, Engagement and Performance." *AI & SOCIETY* 39 (1): 369–388. https://doi.org/10.1007/s00146-022-01514-y.

Rosenberg-Kima, R. B., Y. Koren, and G. Gordon. 2020. "Robot-supported Collaborative Learning (RSCL): Social Robots as Teaching Assistants for Higher Education Small Group Facilitation." *Frontiers in Robotics and AI* 6 (148): 1–12.

Rutatola, E. P., K. Stroeken, and T. Belpaeme. 2025. "'Habari, Colleague!': A Qualitative Exploration of the Perceptions of Primary School Mathematics Teachers in Tanzania regarding the Use of Social Robots." *Applied Sciences* 15 (15): 8483. https://doi.org/10.3390/app15158483.

Šabanović, S., V. Charisi, T. Belpaeme, C. L. Bethel, M. Matarić, R. Murphy, and S. Levy-Tzedek. 2023. "'Robots for Good': Ten Defining Questions." *Science Robotics* 8 (84): eadl4238. https://doi.org/10.1126/scirobotics.adl4238.

Sandygulova, A., W. Johal, Z. Zhexenova, B. Tleubayev, A. Zhanatkyzy, A. Turarova, Z. Telisheva, A. CohenMiller, T. Asselborn, and P. Dillenbourg. 2020. "Cowriting Kazakh: Learning a New Script with a Robot." In *ACM/IEEE International Conference on Human-Robot Interaction, 2016*, 113–120. Cambridge, UK: ACM. https://doi.org/10.1145/3319502.3374813.

Serholt, S. 2018. "Breakdowns in Children's Interactions with a Robotic Tutor: A Longitudinal Study." *Computers in Human Behavior* 81:250–264. https://doi.org/10.1016/j.chb.2017.12.030.

Serholt, S., and W. Barendregt. 2016. "Robots Tutoring Children." In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, 1–10. Gothenburg, Sweden. https://doi.org/10.1145/2971485.2971536.

Serholt, S., W. Barendregt, A. Vasalou, P. Alves-Oliveira, A. Jones, S. Petisca, and A. Paiva. 2017. "The Case of Classroom Robots: Teachers' Deliberations on the Ethical Tensions." *AI & SOCIETY* 32 (4): 613–631. https://doi.org/10.1007/s00146-016-0667-2.

Serholt, S., L. Pareto, S. Ekström, and S. Ljungblad. 2020. "Trouble and Repair in Child–Robot Interaction: A Study of Complex Interactions with a Robot Tutee in a Primary School Classroom." *Frontiers in Robotics and AI* 7:1–13. https://doi.org/10.3389/frobt.2020.00046.

Shamsi, M., M. Zakerijenad, and M. R. Zaraifard. 2025. "Waste to Emerging and Sustainable Wealth: An Integrated Mini 4.0-Recycling 4.0-Decision Making 4.0 Framework Overcoming Greener red mud Recycling Technologies Promotion." *Environmental Technology & Innovation* 40:104440. https://doi.org/10.1016/j.eti.2025.104440.

Smakman, M. 2023. "Robots in Education: Implementing Robot Tutors in a Morally Justified Way." Vrije Universiteit Amsterdam. https://doi.org/10.5463/thesis.133.

Smakman, M., P. Vogt, and E. A. Konijn. 2021. "Moral Considerations on Social Robots in Education: A Multi-stakeholder Perspective." *Computers & Education* 174:104317. https://doi.org/10.1016/j.compedu.2021.104317.

Sonderegger, S. 2022. "How Generative Language Models Can Enhance Interactive Learning with Social Robots." In *Proceedings of the 19th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2022)*. Lisbon, Portugal.

Song, H., E. I. Barakova, J. Ham, and P. Markopoulos. 2024. "The Impact of Social Robots' Presence and Roles on Children's Performance in Musical Instrument Practice." *British Journal of Educational Technology* 55 (3): 1041–1059. https://doi.org/10.1111/bjet.13416.

Song, H., E. I. Barakova, P. Markopoulos, and J. Ham. 2021. "Personalizing HRI in Musical Instrument Practicing: The Influence of Robot Roles (Evaluative versus Nonevaluative) on the Child's Motivation for Children in Different Learning Stages." *Frontiers in Robotics and AI* 8:1–12. https://doi.org/10.3389/frobt.2021.699524.

Song, H., K. Tsiakas, J. Ham, P. Markopoulos, and E. I. Barakova. 2024. "'How Would You Score Yourself?': The Effect of Self-assessment Strategy through Robots on Children's Motivation and Performance in Piano Practice." *International Journal of Social Robotics* 16 (2): 327–340. https://doi.org/10.1007/s12369-023-01080-3.

Spolaôr, N., and F. B. V. Benitti. 2017. "Robotics Applications Grounded in Learning Theories on Tertiary Education: A Systematic Review." *Computers & Education* 112:97–107. https://doi.org/10.1016/j.compedu.2017.05.001.

Strauss, H., and L. H. Zeigler. 1975. "The Delphi Technique and Its Uses in Social Science Research." *The Journal of Creative Behavior* 9 (4): 253–259. https://doi.org/10.1002/j.2162-6057.1975.tb00574.x.

Susskind, L., S. McKearnan, and J. Thomas-Larmer. 1999. *The Consensus Building Handbook : A Comprehensive Guide to Reaching Agreement*. Thousand Oaks, CA: Sage Publications.

Tanaka, F., and S. Matsuzoe. 2012. "Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning." *Journal of Human-Robot Interaction* 1 (1): 78–95. https://doi.org/10.5898/JHRI.1.1.Tanaka.

Tolksdorf, N. F., S. Siebert, I. Zorn, I. Horwath, and K. J. Rohlfing. 2021. "Ethical Considerations of Applying Robots in Kindergarten Settings: Towards an Approach from a Macroperspective." *International Journal of Social Robotics* 13 (2): 129–140. https://doi.org/10.1007/s12369-020-00622-3.

Topsakal, O., and E. Topsakal. 2022. "Framework for a Foreign Language Teaching Software for Children Utilizing AR, Voicebots and ChatGPT (Large Language Models)." *The Journal of Cognitive Systems* 7 (2): 33–38. https://doi.org/10.52876/jcs.1227392.

Tozadore, D., N. Ertug, Y. Chaker, and M. Abderrahim. 2025. "RoboBuddy in the Classroom: Exploring LLM-Powered Social Robots for Storytelling in Learning and Integration." In *The Proceedings of IEEE RO-MAN 2025*, 1543–1549. Eindhoven, The Netherlands. https://doi.org/10.1109/RO-MAN63969.2025.11217800.

Tozadore, D. C., C. Wang, G. Marchesi, B. Bruno, and P. Dillenbourg. 2022. "A Game-Based Approach for Evaluating and Customizing Handwriting Training Using an Autonomous Social Robot." In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1467–1473. Naples, Italy. https://doi.org/10.1109/RO-MAN53752.2022.9900661.

van den Berk-Smeekens, I., M. W. P. de Korte, M. van Dongen-Boomsma, I. J. Oosterling, J. C. den Boer, E. I. Barakova, T. Lourens, J. C. Glennon, W. G. Staal, and J. K. Buitelaar. 2022. "Pivotal Response Treatment with and without Robot-Assistance for Children with Autism: A Randomized Controlled Trial." *European Child & Adolescent Psychiatry* 31 (12): 1871–1883. https://doi.org/10.1007/s00787-021-01804-8.

Van Lehn, K. 2011. "The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems." *Educational Psychologist* 46 (4): 197–221. https://doi.org/10.1080/00461520.2011.611369.

van Minkelen, P., C. Gruson, P. van Hees, M. Willems, J. de Wit, R. Aarts, J. Denissen, and P. Vogt. 2020. "Using Self-determination Theory in Social Robots to Increase Motivation in L2 Word Learning." In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 369–377. Cambridge, UK. https://doi.org/10.1145/3319502.3374828.

van Otterdijk, M., M. de Korte, I. van den Berk-Smeekens, J. Hendrix, M. van Dongen-Boomsma, J. den Boer, J. Buitelaar, et al. 2020. "The Effects of Long-Term Child–Robot Interaction on the Attention and the Engagement of Children with Autism." *Robotics* 9 (4): 79. https://doi.org/10.3390/robotics9040079.

Vemprala, S. H., R. Bonatti, A. Bucker, and A. Kapoor. 2024. "ChatGPT for Robotics: Design Principles and Model Abilities." *IEEE Access* 12:55682–55696. https://doi.org/10.1109/ACCESS.2024.3387941.

Vogt, P., M. de Haas, C. de Jong, P. Baxter, and E. Krahmer. 2017. "Child-Robot Interactions for Second Language Tutoring to Preschool Children." *Frontiers in Human Neuroscience* 11:1–7. https://doi.org/10.3389/fnhum.2017.00073.

Vogt, P., R. van den Berghe, M. de Haas, L. Hoffman, J. Kanero, E. Mamus, J.-M. Montanier, and A. K. Pandey. 2019. "Second Language Tutoring Using Social Robots: A Large-Scale Study." In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 497–505. Daegu, South Korea. https://doi.org/10.1109/HRI.2019.8673077.

Voultsiou, E., E. Vrochidou, L. Moussiades, and G. A. Papakostas. 2025. "The Potential of Large Language Models for Social Robots in Special Education." *Progress in Artificial Intelligence* 14:165–189. https://doi.org/10.1007/s13748-025-00363-2.

Wang, F., and A. C. K. Cheung. 2025. "Robots' Social Behaviors for Language Learning: A Systematic Review and Meta-analysis." *Review of Educational Research* 95 (2): 175–212. https://doi.org/10.3102/00346543231216437.

Wang, K., G.-Y. Sang, L.-Z. Huang, S.-H. Li, and J.-W. Guo. 2023. "The Effectiveness of Educational Robots in Improving Learning Outcomes: A Meta-analysis." *Sustainability* 15 (5): 4637. https://doi.org/10.3390/su15054637.

Wang, Q., T. Anikina, N. Feldhus, J. Genabith, L. Hennig, and S. Möller. 2024. "LLMCheckup: Conversational Examination of Large Language Models via Interpretability Tools and Self-explanations." In *Proceedings of the Third Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, 89–104. Mexico City, Mexico. https://doi.org/10.18653/v1/2024.hcinlp-1.9.

Westlund, J. M. K., H. W. Park, R. Williams, and C. Breazeal. 2018. "Measuring Young Children's Long-Term Relationships with Social Robots." In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 207–218. Trondheim, Norway. https://doi.org/10.1145/3202185.3202732.

Woo, H., G. K. LeTendre, T. Pham-Shouse, and Y. Xiong. 2021. "The Use of Social Robots in Classrooms: A Review of Field-Based Studies." *Educational Research Review* 33:100388. https://doi.org/10.1016/j.edurev.2021.100388.

Yadollahi, E., W. Johal, A. Paiva, and P. Dillenbourg. 2018. "When Deictic Gestures in a Robot Can Harm Child-Robot Collaboration." In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 195–206. Trondheim, Norway. https://doi.org/10.1145/3202185.3202743.

Youssef, K., S. Said, S. Alkork, and T. Beyrouthy. 2023. "Social Robotics in Education: A Survey on Recent Studies and Applications." *International Journal of Emerging Technologies in Learning (IJET)* 18 (3): 67–82. https://doi.org/10.3991/ijet.v18i03.33529.

Zhang, C., J. Chen, J. Li, Y. Peng, and Z. Mao. 2023. "Large Language Models for Human–Robot Interaction: A Review." *Biomimetic Intelligence and Robotics* 3 (4): 100131. https://doi.org/10.1016/j.birob.2023.100131.

Zhao, Z., D. Zheng, Y. Chen, J. Luo, Y. Wang, P. Huang, and C. Yang. 2025. "Integrating with Multimodal Information for Enhancing Robotic Grasping with Vision-Language Models." *IEEE Transactions on Automation Science and Engineering* 22:13073–13086. https://doi.org/10.1109/TASE.2025.3550360.

Zou, J., S. Gauthier, H. Pellerin, T. Gargot, D. Archambault, M. Chetouani, D. Cohen, and S. M. Anzalone. 2024. "R2C3, a Rehabilitation Robotic Companion for Children and Caregivers: The Collaborative Design of a Social Robot for Children with Neurodevelopmental Disorders." *International Journal of Social Robotics* 16 (3): 599–617. https://doi.org/10.1007/s12369-024-01104-6.