Progress in Artificial Intelligence (2025) 14:165–189
https://doi.org/10.1007/s13748-025-00363-2

REVIEW

The potential of Large Language Models for social robots in special education

Evdokia Voultsiou[1] · Eleni Vrochidou[2] · Lefteris Moussiades[1] · George A. Papakostas[2]

Received: 17 April 2024 / Accepted: 4 January 2025 / Published online: 6 February 2025

Abstract

Large language models (LLMs) have created remarkable possibilities for analyzing and generating language data and have been integrated into several fields aiming to transform them, including education. While most research efforts focus on LLMs in typical education or social robots, limited applications of LLMs have been reported in special education. Moreover, there is a profound lack of combined research in LLM-based social robots in special education. In this work, we argue that although LLMs and social robots have demonstrated their potential to advance special education separately, their combination is not yet fully exploited, and further research is required to enable such use. The first objective of this work is to review relevant literature to assess the feasibility of developing LLMs on social robot platforms for use in special education. The second objective of this work is to reveal related challenges, limitations, opportunities, and ethical considerations to provide insights, aiming to subsequently formulate guidelines for the efficient integration of LLM-based social robots into special education practices. To this end, the third objective of this work is to propose a thoughtful framework, aiming to formulate a safe and inclusive learning environment for students in special education, suggesting actionable steps that could be followed by educators, developers and stakeholders, towards address the unique needs and challenges of students with diverse learning requirements.

Keywords Large Language Models · Social robots · Special education · Artificial intelligence · Educational technologies

1 Introduction

Large Language Models (LLMs) are artificial intelligence (AI) algorithms based on transformer models, a type of deep neural network. They incorporate billions of parameters and are pre-trained with vast language data to learn underlying patterns and language rules. The latter enables them to understand and generate original content [1].

Recently, LLMs are playing a leading role in a wide range of natural language processing (NLP) applications involving language generation [2], automatic text summarization [3], text comprehension [4] and classification [5]. The most up-to-date LLMs, such as the Generative Pre-trained Transformer 4 (GPT-4) [6] and Large Language Model Meta AI (LLaMA) [7], have proven their capability to comprehend and generate efficiently human-resembling text. Thus, they are adopted as powerful tools for multiple applications, in public health and education. Therefore, the use of LLMs has been investigated for applications in the field of mental health as supportive tools [8] in the medical/clinical field: for diagnosis of mental distress [9], for cognitive impairments such as Alzheimer's disease [10], to predict the mini-mental state examination score related to cognitive impairments [11], for ASD detection [12] and more.

In the educational context, the use of LLMs deems ambitious; early applications reveal the potential of LLMs in the

✉ George A. Papakostas
gpapak@cs.duth.gr

Evdokia Voultsiou
evoultsi@cs.duth.gr

Eleni Vrochidou
evrochid@cs.duth.gr

Lefteris Moussiades
lmous@cs.duth.gr

[1] Department of Informatics, School of Science, Democritus University of Thrace, Kavala 65404, Greece

[2] MLV Research Group, Department of Informatics, Democritus University of Thrace, Kavala 65404, Greece

educational teaching/learning process for knowledge tracing [13], socio-emotional support of online conventional agents [14], production of learning resources [15], as well as for several different subjects such as programming [15], mathematics [16, 17], science [18], and medicine [19].

Social robots have also been used in education as teaching assistants, tutors, or peers [20]. Their implementation has reported increased academic and cognitive outcomes, while their acceptance from teachers, students, and parents is nowadays undoubtable [21]. The benefits of the use of social robots have been extended to special education, towards improving the cognitive, emotional, and social development of children with certain impairments, such as autism spectrum disorder (ASR), hearing impairments, down syndrome, neuro-developmental disorder, cerebral palsy, and more [22].

Based on the above, the empowerment of education using LLMs, combined with social robots, could bring out enhanced possibilities. On the one hand, LLMs have been integrated into robotics towards intelligent interactions and fulfilled autonomy for perception, control, decision-making, and path planning [23], mainly for industrial applications [23, 24]. On the other hand, integrating LLMs in social robots for education, specifically for special education, is still in its infancy.

To this end, the first main objective of this work is to identify the potential of LLMs for social robots in special education. The scope is to bring together all related research in LLM-based social robots in special education and investigate all referenced implementations, highlighting challenges and identifying opportunities for their efficient integration into special education practices. It should be noted that, to the authors' knowledge, there is no similar review article up to date focusing solely on LLMs for social robots in special education.

In 2021–2022, 7.3 million students between the ages 3 and 21 received special education and related services under the Individuals with Disabilities Education Act (IDEA) in the United States, equal to a percentage of 15% of all public-school students [25]. Based on the same source, the four most common disability types are specific language disabilities (32%), speech or language impairments (19%), health impairments (15%), and autism (12%).

According to the United Nations Educational, Scientific and Cultural Organization (UNESCO), special education is a general framework of learning strategies that must be adjusted to respond to these educational needs [26]. According to the United Nations, based on Article 24 of the Convention on the Rights of Persons with Disabilities (CRPD), there is a rightful claim for every person with mental or physical impairments to be educated. Individuals with special needs should not be excluded from society, and any action should be undertaken so as for their mental and emotional state to be sustained at an optimal level [27].

United Nations International Children's Emergency Fund (UNICEF) in 2022 acknowledged the significance of Assistive Technologies (ATs) for children with neurodevelopment disorders, and Social Robots are included among their recommendations [28]. ATs bolster students with special educational needs (SEN) to alleviate any disabilities regarding their senses or mentality so that they can be socially active and receptive to acquiring various skills [29]. Furthermore, AI in special education is considered to be another assistive educational tool for integration, triggering positive impacts [30]. SEN students can achieve their learning goals, as they relish general freedom in the learning process, through flexible and adaptive personal tutoring. AI technologies encompass various application sectors, especially LLMs, and reinforce the learning process in special education [31].

Based on the above, the importance and ever-increasing need for special education initiatives and the impact of both LLMs and social robots in the learning process are evident. Therefore, the combination of LLMs and social robots has the potential to revolutionize special education.

Despite the potential of LLM-based social robots in special education, there is no relevant research to include field trials, current applications, their potential impact, limitations, challenges, and related ethical concerns. Moreover, while little is known about the impact of LLMs in education, nothing is reported regarding the impact of LLM-based social robots in either typical or special education, regarding students' motivation, engagement, learning outcomes, and more.

Reviews on LLMs for human-robot interaction [23], on robot-assisted special education [32], on social robots for special education [22], on LLMs for typical education [33], have been conducted in the literature. However, their combination has not previously been investigated. A preliminary discussion of AI as a technology with the potential to significantly change special education practices been published recently [34], yet it deals with AI software and future considerations and does not report implementations of LLMs on social robots for special education. Therefore, this work constitutes the first approach to advance the knowledge and understanding of LLMs' current role and potential for social robots in special education. Apart from the systematic report of all relevant research in the field, the second main objective of this work is to contribute to the corpus of knowledge by highlighting related challenges and opportunities and providing insights into how LLM-based social robots can be effectively integrated into special education practices and fundamentally promote them. Towards this direction, the third main objective of this work is to propose a framework for the integration of LLM-based social robots in special education.

Based on the three main objectives of this works, three research questions (RQ) have been formulated to structure and guide the conducted research:

1. RQ1: “What is the current status of LLM-based social robots in special education?”.
2. RQ2: “What are the challenges, limitations and considerations of applying LLM-based social robots in special education?”.
3. RQ3: “It is possible to formulate a framework for LLM-based social robot integration in special education?”.

In conclusion, the contributions of this work can be summarized as follows:

- This work constitutes a systematic report dealing with the current role and potential of LLMs for social robots in special education, which has not previously been reported in the literature.
- This work underlines limitations, challenges, and ethical concerns regarding the integration of LLM-based robots in special education.
- This work investigates for the first time the potential impact of LLM-based social robots in special education, regarding students’ motivation, engagement, learning outcomes, and more.
- This work, based on the gathered evidence, proposes the first general framework for the integration of LLM-based social robots in special education, providing actionable steps that could be taken by educators, developers and stakeholders.

The rest of this work is structured as follows: Sect. 2 includes a brief overview of LLMs and discusses the use of social robots in special education. The research methodology followed in this work is presented in Sect. 3, while Sect. 4 investigates LLM-based social robots in special education, including the investigation of LLMs for social robots and LLMs in special education separately. Section 5 discusses research findings, limitations, challenges and ethical issues for the integration of LLM-based social robots in special education, while Sect. 6 introduces a framework for the integration of LLM-based social robots in special education. Finally, Sect. 7 concludes the paper.

2 Background

Large Language Models (LLMs) and social robotics independently demonstrated transformative potential across various domains, including education, healthcare, and social interactions. LLMs, with their ability to understand and generate human-like language, provide adaptive conversational tools that cater to individual needs. Social robots, on the other hand, engage users through physical presence and interactivity, often serving as assistants, companions, or therapeutic agents. In special education, where students usually require tailored and responsive support, integrating these two technologies opens unique avenues for enhancing learning and social engagement.

Combining LLMs with social robots creates a powerful synergy where the conversational intelligence of LLMs complements the interactive embodiment of social robots. This combination is promising for special education, where diverse learning requirements require adaptive, emotionally sensitive, and personalized interventions. This section reviews the foundational aspects of LLMs and social robots and sets the stage for understanding their joint application. We highlight how each technology contributes distinctively to supporting students with special needs. This background is crucial to appreciate the challenges, opportunities, and ethical considerations in deploying LLM-based social robots in special education contexts.

2.1 LLMs - overview

LLMs are machine learning models with large architectures, able to generate context full of coherence and accuracy, replicating human speech by calculating the probability of a word following a certain input. LLMs can distinguish patterns in words and predict each following word after being trained with large language datasets, namely corpus. Pretrained LLMs are fine-tuned to find practical use in tasks such as translation, summarization, domain-specific knowledge generation, etc [35].

The foundations of LLMs can be tracked in the 40s, when McCulloch and Pitts [36] introduced the concept of artificial neural networks (ANNs), while in the 50s, the first rule-based language model was presented by IBM-Georgetown University who developed a Russian-English translation system [37]. Another important development was the first chatbot, ELIZA, launched in the 60s [38]. Eliza was the earliest example of a language model; it was based on rules and pattern-matching techniques. Although a simple model, Eliza could identify keywords from the user input and match a pre-programmed answer. This was a significant milestone in the development of language models, as it demonstrated the potential of natural language processing and paved the way for more sophisticated models to come. In the 90s, the introduction of Long Short-Term Memory (LSTM) [39] revealed new opportunities for developing deeper neural networks able to capture statistical patterns of larger amounts of data, aiming to create statistics-based language models. In the 2010s, recurrent neural language models (RNNLM) were introduced, generating more natural texts than previous approaches [40].

At the same time, Stanford's CoreNPL suite was launched [41], enabling sentiment analysis, along with GoogleBrain [42] which provided word embeddings towards clarifying text comprehension. All previously mentioned enhancements contributed to developing Google's Neural Machine Translation System [43]. However, the burst of LLMs was marked after the development of transformer models, introduced in 2017 [44]. Based on transformer architectures, in 2018, the Generative Pre-trained Transformer (GPT) model [45] and Bidirectional Encoder Representations from Transformers (BERT) [46] were developed. The next version of GPT, GPT-2 [47], used unsupervised pre-trained models for supervised tasks towards multi-tasking learning while training without fine-tuning. OpenAI's GPT-2 is regarded as the first LLM. At that time, other transformer-based LLMs were also developed, such as the Megatron-LM [48]. In 2020, GTP-3 was released [1], including more advanced features in answering questions, translation and searching, and being able to generate more natural language output with less fine-tuning, forming the basis of ChatGPT [49, 50].The latest version, GTP-4 [6], offered even more opportunities to analyze nonverbal data and generate even more realistic textual output.

Indicatively, some of the most recent LLM models launched in 2023 are the following:

- **GPT-4**: Released in March 2023 by OpenAI, it is the most updated version of the GPT series and is used mainly to generate human-like language [6].
- **PaLM 2**: Released in May 2023 by Google, AI empowers the chatbot of Google, Bard [51].
- **LLaMa-2**: Released in July 2023 by Meta AI and Micrososft, it is free to use. LLaMa 2– Chat is a model for dialogues [52].
- **Falcon**: Released in September 2023 by the Technology Innovation Institute [53], it owns three variants depending on the number of its parameters.

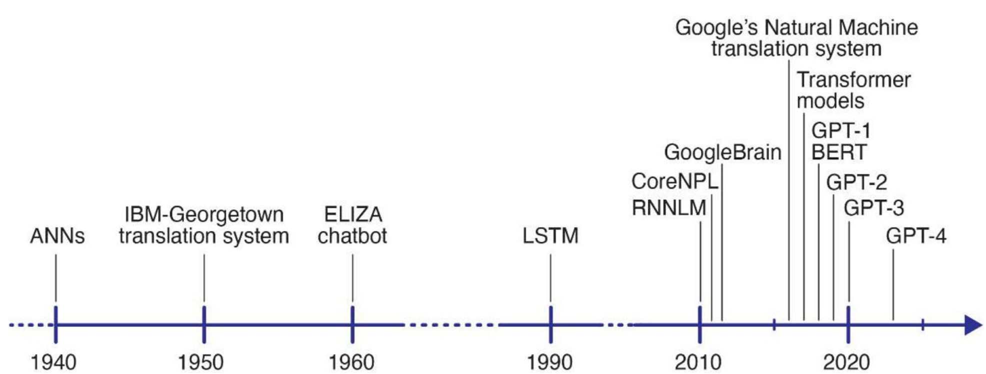
Figure 1 depicts the milestones in the history of LLMs as considered by the authors.

2.2 Social robots in special education (SE & SR)

Social robots are robots that can interact with humans in a socially acceptable way [22]. Special education refers to the educational services provided to students with disabilities. Special education, among others, differs from the typical one in the sense that it does not follow a common curriculum for all students of the same class. Special education needs to be designed to meet the special needs of each different impairment and, furthermore, the personal needs of each individual student. Social robots can be highly captivating and can motivate students to try harder on tasks that otherwise would refuse to undertake due to their impairment [54]. Moreover, AI capabilities denoted to social robots allowed for efficient and adaptive human-robot interactions. Social robots offer adequate 'safety' to students to try practicing skills without the fear of judgment or criticism, thus reducing their anxiety and focus on the goals of each educational task [55].

Social robots gave insight into social-based engagement in education and, in particular, met a high demand in the field of Special Educational Needs and Disabilities [56]. Every student possesses a palette of heterogeneous emotional and behavioral characteristics in the learning process within the classroom, especially the vulnerable ones struggling with physical and mental issues [57]. Adaptive learning strategies to the uniqueness of each individual are paramount [22, 58]. Social robots have proven their effectiveness as therapeutic tools in special education towards promoting social, cognitive, and intellectual skills, assigned in different roles: teacher, assistant, or peer [20], as well as for

Fig. 1 Milestones in the history of LLMs



several different impairments, ASD, mobility issues, cerebral palsy, attention deficit hyperactivity disorder (ADHD), hearing impairments, Down syndrome, oncological disorders, neuro-developmental disorders (NDD) [22]. It should be noted that the use of social robots in special education is not the focus of this work; it is a subject that has been extensively researched in the literature, and as such, does not need to be systematically reviewed again in the context of this paper. Yet, it is useful to discuss how LLMs are able to contribute to them.

Considering that LLMs are capable of generating adaptive and interactive conversations, it is obvious that their integration could enhance the verbal interactivity of a social robot. Moreover, emphatic AI [59], which is set upon the mutual exchange of emotions in human-robot interaction, could endow social robots with adequate emotional content so that any input call can be used to adjust its reaction by mimicking human feelings. An educator in the form of a social robot, with the aid of a Generative Pretrained Transformer is advocated [60]. To this end, GPT-3 has been proven to achieve "depth and complexity" [61] in learning procedures through speech, e.g., conversation, complex questions, and more. According to Bhat et al. [61], whether the user or the robot has either a passive or an active role, GPT-3 provides a multi-layered cognitive approach, as the model can readily adjust to the user's demands. The general perception of freedom of speech through a social robot forms prosperous dynamics in the scientific community.

3 Research methodology

The conducted research methodology was based on the three research questions (RQ) aligned with the main objectives of this works according to the guidelines provided by Kitchenham [62] to conduct systematic literature reviews:

1. RQ1: "What is the current status of LLM-based social robots in special education?".
2. RQ2: "What are the challenges, limitations and considerations of applying LLM-based social robots in special education?".
3. RQ3: "It is possible to formulate a framework for LLM-based social robot integration in special education?".

These three research questions are used as the backbone to structure this work. Therefore, each question guides a specific section of the manuscript, as seen in Table 1, ensuring cohesion with the main objectives, as well as a clear and narrative flow.

In this work, the PRISMA statement [63] was followed to conduct a clear, transparent and comprehensive systematic review. The PRISMA diagram of the conducted research methodology is illustrated in Fig. 2.

The original research was conducted in the Scopus database. Scopus is considered as the most comprehensive and authentic database of scholarly publications since it indexes only curated content of high quality that is annually re-evaluated by an advisory board [64]. The same research terms were also used in Google Scholar database so as to enhance the number of retrieved articles.

At the first step of the research methodology, four targeted queries (Q) were executed within the article title, abstract, and keywords, focusing to identify (Identification) the relevant literature on LLMs for social robots in special education:

1. Q1: "Special Education" AND "Social Robots" (SE & SR): returned 40 documents. Since our research needs to include LLMs, the latter query is secondary and it is needed to structure the background of social robots in special education, therefore only review articles from the results were considered, limiting the documents to two.
2. Q2: "Large Language Models" AND "Social Robots" (LLM &SR): returned 21 documents.
3. Q3: "Large Language Models" AND "Special Education" (LLM &SE): returned zero documents. Yet, the

Table 1 Structure of the paper based on the research questions (RQ), and overview of the final set of publications for each targeted query (Q) per database based on the defined taxonomy

RQ	Q	Taxonomy	Number of publications	Scopus Refs.	Scholar Refs.	Section; subsection
Background on social robots in special education						Section 2;
	1	SE & SR	2	[22, 54]	-	2.2
	LLMs in social robots					**Section 4;**
	2	LLM & SR	16	[65–74]	[75–80]	4.1
		LLMs in special education				**Section 4;**
	3	LLM& SE	5	[81–85]	[86]	4.2
RQ1	**LLMs for social robots in special education**					**Section 4;**
	4	LLM & SR & SE	2	-	[87, 88]	4.3
RQ2	**Related challenges, limitations and considerations**					**Section 5**
RQ3	**Framework for the integration of LLM-based social robots in special education**					**Section 6**

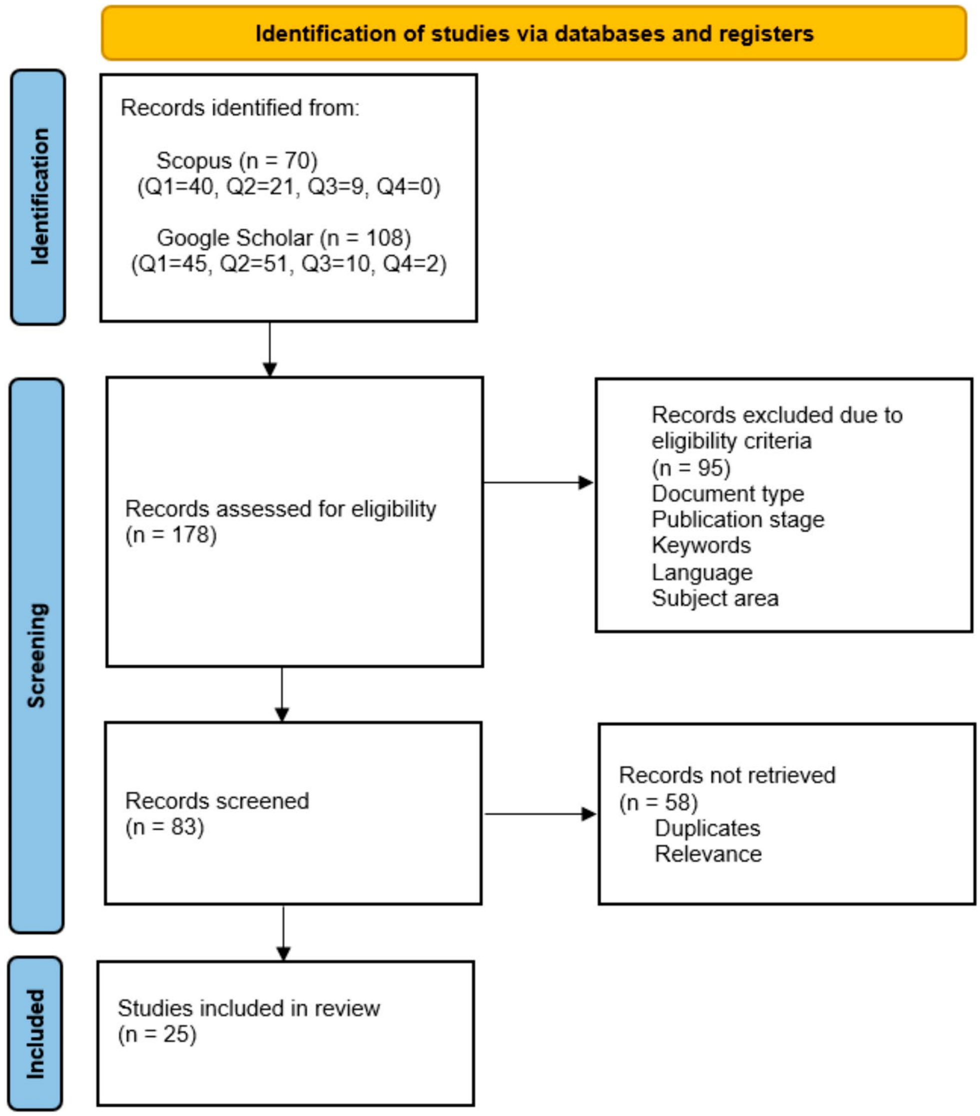


Fig. 2 The PRISMA diagram of the applied research methodology

broader term of "special education" seems to limit the results, while by using the term "Education" instead, research returned 818 documents. Therefore, special education approaches were located withing the bibliography by using more refine searching rules by naming the most common types of special education categories, (e.g., learning disabilities, language disabilities such as dyslexia, and autism). Query of "Large Language Models" AND "autism" returned six documents, "dyslexia" returned one, "learning disabilities" returned one, "language disabilities" returned zero, and "speech impairment" returned one. In total, the latter approach returned 9 documents. Note that only applications in special education were included, therefore LLM-based diagnostic tools for impairments detection were not included, as consider related more to healthcare rather than special education.

4. Q4: "Large Language Models" AND "Social Robots" AND "Special Education" (LLM & SR & SE) returned zero documents.

The fourth query (Q4) comes as an inclusion of all the previous ones. However, even though LLMs have been incorporated into social robots and used in special education, conjoint research on LLM-based social robots in special education is scarce in the literature, indicating an uncharted research area. It should be noted that LLMs have caught research attention, especially GPT-based models, after 2019. Potential application fields and limitations have not yet been fully depicted; therefore, any further integration in specialized domains is still in its infancy. During this first step of the research methodology, 178 documents in total were located.

During the second step of the methodology (Screening), the following eligibility criteria were applied:

(a) The document types must be either book chapters, journals, or conference papers (For Q1 only review articles).
(b) The literature must be at the final publication stage.
(c) The keywords are limited to the words specified for each query.
(d) The text must be written in English language.
(e) The papers must belong to the subject area of Computer Science or Engineering.

Eligibility criteria concluded in 83 documents. Finally, at the third step of the methodology (Included), information extraction from the abstracts and main text of the documents and classification of the literature retrieved from the previous steps took place to delete duplicates between the two databases, include only the publications related to the subject focusing on applied works, and classify them based on the search queries, concluding to 25 documents.

All defined works from the conducted research are referenced within this article, either in the text or in pivot tables, and are equally used to draw conclusions.

To this end, a taxonomy is defined, to categorize the 25 gathered documents systematically, so as to provide a clear framework enabling easier navigation to their content, facilitating their comprehensive analysis, towards making it easier to identify patterns, reveal trends, gaps and relationships among them.

The proposed taxonomy aims to classify and organize the documents into four hierarchical categories based on their content, as imposed from the conducted research, i.e., the corresponding queries (Q):

1. SE & SR: Documents reporting the use of social robots in special education.
2. LLM & SR: Documents reporting the use of LLMs in social robots (SR).
3. LLM& SE: Documents reporting the use of LLMs in special education (SE).
4. LLM & SR & SE: Documents reporting the combination of LLMs and social robots in special education.

The 25 selected papers are listed in Table 1 and grouped based on the defined taxonomy. Note that the documents retrieved regarding the use of social robots in special education (SE & SR) are used to define the backgrounds provided in Sect. 2.2.

The rest of the paper is structured based on the research questions. Therefore, each of the following sections aims to answer one research question: Sect. 4 for RQ1, Sect. 5 for RQ2, and Sect. 6 for RQ3. For the structure of Sect. 4, where the aim is to review the current status of LLM-based social robots in special education, the defined taxonomy is used to organize the examination of the gathered documents in three distinct subsections (4.1, 4.2 and 4.3), as summarized in Table 1.

4 Review of the literature

Early research on the integration of LLMs in education showed that it could meliorate both teaching and learning experiences [65]. The learning opportunities seem endless, according to Kasneci et al. [66]. The authors claim that LLMs could be integrated into education for all levels of education, as well as for professional development. For learning tasks, LLMs could be used (1) in elementary schools to support students in practicing writing and reading by providing corrections, to encourage students on critical thinking, to summarize or interpret information for them for reading comprehension, to generate questions to organize their study, (2) in high schools, to additionally help on analytical thinking, and problem-solving, to generate exercises for practice for a variety of high-school curriculum subjects, (3) in universities [66], to support research by providing valuable resources on highly specialized scientific topics, (4) from remote learners, to guide turn-taking during conferences, to engage participants, (5) from professional learners, to generate domain-specific knowledge, (6) from learners with special needs, to empower them with abilities they lack [67, 68].

For teaching tasks, LLMs could be used by teachers for (1) personalized teaching, to adapt lessons to each individual student's need, (2) for creating teaching material that could be diverse, targeted, and of all levels of difficulties, (3) for reading and writing tasks, as to summarize texts and highlighting main points to better deliver the lesson to students, (4) for evaluation of students, to help them correct essays regarding grammar and spelling issues, to check reports for plagiarism [66, 69, 70].

Therefore, the use of LLMs in typical education mainly focuses on personalized learning to individual students' needs, on educational content generation used from both students and teachers, as well as for assessment, i.e., grading, and instant feedback. In special education, the use of LLMs remains in the same context. Yet, a more specialized use is intended, focusing more on improving the accessibility for students with disabilities, e.g., with content generation in accessible formats, such as Braille, audio, or text, for students with specific impairments, dyslexia, non-verbal students, etc. Moreover, for students in need of personalized behavioral and emotional support, such as in autism, LLMs can provide adequate support to manage their learning environment more effectively, by providing highly customized learning plans. In terms of technical and implementation aspects, LLMs in typical and special education differ in the used datasets; in typical education the datasets are designed to cover a wide range of subjects and educational levels, while in special education the datasets content is more specialized, focusing on each different underlying impairment.

Overall, the potential of LLMs is anticipated at all stages of education for both teachers and students, as well as in industry to enhance the control process of soft robotics [71–73]. While many research articles discuss the potential of LLMs in education [74–77], there is no solid work to deal with their prospects in special education. Furthermore, while the use of social robots in special education has been at the forefront for many years, there is limited research on the use of LLM-based social robots in special education.

In what follows, the integration of LLMs in social robots, the use of LLMs in special education, and the combination of the latter two in LLM-based social robots for special education are exhaustively investigated, following a structure for this section as imposed by the proposed taxonomy, aiming to provide answers to RQ1.

This section aims to present the current status so as to conclude the stemming opportunities of such a combination and present all related challenges towards providing insight and guidelines for the use of LLM-based social robots towards their efficient integration into special education practices. It should be noted here that most of the referenced works in the following, do not report numerical evaluation results, as they mainly simulate proposed frameworks to underscore the transformative potential of LLMs for robots and in educational applications. The system evaluation is mostly conducted by subjective questionnaires from the participants. Thus, in cases where technical aspects are reported, such as datasets and performance results, the latter are referred in the main text and not included in the commutative Tables due to their limited number. However, the used LLM model for all cases is referenced in the Tables.

4.1 LLMs for social robots (LLM & SR)

The evolution of socially interactive robots and their integration into human daily lives have been supported by simultaneous enhancements from other related scientific fields towards even increased human-robot interaction. Zhang et al. [78] conducted an extended review on the advancements of LLMs in human-robot interaction based on the recent progress in the field, aiming to provide directions for future research.

To this end, virtual reality (VR) has been coupled with LLMs and social robots [79] to deliver an immersive interactive English language teaching experience. The integration of LLMs in human-robot interaction has previously shown promising results; Ye et al. [80] proposed a ChatGPT-based assistant robotic arm (Franka Emika Panda robot arm) and concluded that such an integration increased the trust in human-robot collaboration due to the more efficient acquired communication skills of the robotic arm. The latter motivated Bottega et al. to adapt LLMs to their language learning games, integrating the GPT-4 model into the virtual robot of their VR application. Results indicated adaptability and quick feedback from the robot, which facilitated the interaction and engaged the users (no numerical performance evaluation results are reported).

Murali et al. [81] introduced a framework that used ChatGPT to develop a group-facilitation social robot. Dyads of participants interacted with the social robot to select the best candidate out of six fictional resumes for a manager's position. The authors' findings include a high percentage of speaker label identification (77% accuracy from transcribed tecta and 90% word level accuracy), indicating the potential of LLMs as a diarization tool for future systems.

Billing et al. [82] presented the first integration of OpenAI GPT-3 with Pepper and Nao social robots. The authors facilitated an open verbal dialogue with the robots, sharing the potential of using LLM-based social robots in multiple dialogue systems. The technical implementation integrates three different services that constitute the complete dialogue system. No datasets or performance evaluation results are reported. Based on the latter original idea, Axelsson and Skantze [83] developed an application of an LLM-based social robot as a presenter, e.g., as a museum guide. The authors introduced an original approach for lexicalization, i.e., transforming semi-logical representations of chosen language statements from a knowledge graph into natural language. A feedback classifier was also adopted to collect data, i.e., users' multimodal feedback, from the presentation. The feedback was classified as positive, negative, or neutral for updating the grounding status in the knowledge graph accordingly, thus affecting the procedure of the presentation. The evaluation of the system by 43 participants who interacted with it showed that LLM-based robot presenters are considered more human-resembling and flexible compared to the same static implementation of the system, which does not consider the users' feedback. The evaluation process was conducted by a multiple-choice questionnaire.

Adaptability to users' feedback is critical for establishing an effective human-robot interaction. In [84], the authors use LLMs for real-time emotion generation in a human-robot dialogue. More specifically, they used GPT-3.5 for predicting the emotion of a robot's turn in real-time, exploiting the history of the ongoing dialogue, and the robot gestured the predicted emotion with facial expressions. For the system evaluation, the authors collected subjective questionnaire data. To evaluate the prediction capabilities of the model, a prediction confusion matrix was calculated for all emotions using predicted and actual image labels, reporting best performance for "surprise" of 65% and lowest for "anger" with 41%. Overall, results showed the ability of the model to efficiently generate emotions in real-time, which is critical in LLM-based social robot applications where emotional interaction matters, as it is in companionship, therapy, special education, or even in customer service.

Lozano et al. [85] examined the ability of a proposed LLM framework implemented on social robot EVA to assume nonverbal cues by the user. More specifically, the proposed framework included object recognition capabilities and an LLM to propose meals to cook based on the detected ingredients. Nonverbal communication is crucial in cases where words are either absent or not enough to obtain valuable input from nonverbal cues such as gestures, posture, and eye gaze. The authors conducted two scenarios to illustrate the use of their proposed framework, while no performance results were reported.

In [86], user data from Twitter social media accounts are considered to engage users in generated personalized dialogues with an LLM-based social robot Mini. First, the robot uses a summarization LLM to present an overview of the news, and then a Long-Form Question-Answer model (LFQA) to generate related questions. The usability of the robot was evaluated by 17 participants who freely interacted with the robot. More specifically, the evaluation referred to the usability of the proposed conversation skill to interact with the robot by using a questionnaire of 5-point Likert scale. Results indicated both positive (expressiveness of the robot, diversity of topics, updated information) and negative aspects (delays and processing time). It should be noted that Mini is a social robot designed to assist the elderly with mild cognitive impairments. Therefore, the proposed system is designated to be applied to social robots perceived as companions for the elderly. The same robot was employed in [87] to demonstrate the generation of diverse speech that could dynamically adapt to different user profiles. Paraphrasing was also used to prevent dialogues from turning repetitive and monotonous. The evaluation of the user-adapted semantic description generation reported 4.87 s response time for the entire pipeline, while for the models used for paraphrase generation the inference time ranged between 0.88 and 4.04 s, which is encouraging in both cases and shows their great potential when applied to social robots.

Irfan et al. [88] also used a Furhat robot and an LLM to derive multi-modal open-subject dialogues between the robot and senior users. The conducted workshop included the evaluation of 28 elderly participants. The evaluation of the system was conducted through a pre- and post- interaction recorded audio interview with the participants, as well as from the video data analysis of the participants' interaction with the robot. Results indicated smooth and varied in topics conversations, while many challenged were also reported (included in Table 2). Khoo et al. [89] conducted a similar research study by using the social robot QTrobot and 12 senior participants, aiming to provide personalized interactions and thus improve the user experience. The system was evaluated with written surveys and observations, indication the need to improve users' experience through personalized interaction (no numerical evaluations were reported). Wang et al. [90] proposed a framework to generate conversation responses with expressive robot behavior, involving robot Haru, directly from an LLM. The system evaluation was done through a pilot study with 12 participants answering a short free-text experiential survey. Results indicated hallucinations and repetitions, as well as naturalness, entertainment, helpfulness and empathy from the robot's side. No numerical evaluation results were reported.

Jokinen et al. [91] employed LLMs to make Furhat robot chat about culinary delights. Example queries and responses of the robot were reported to demonstrate the linguistic versatility and feasibility of the system, while no evaluation results were reported. In [92], robot Pepper was powered by a dialogue system based on GPT-3 to produce responses to 31 participants' verbal inputs. The system was evaluated through questionnaires. Results revealed high expectations from the robot, strongly connected to human-human interaction. Borg et al. [93] employed Furhat as a virtual patient to create a platform for clinical reasoning in rheumatology. The platform was evaluated by 15 medical students, compared to a semi-linear virtual patient platform, by evaluating the self-perceived accrual of clinical reasoning skills. Results revealed the preference of the students for the robot platform in terms of learning effect and authenticity. In [94], a dialogue system based on LLMs was embodied in a social robot. The system was an ongoing work, and no application and evaluation results were yet reported. Kim et al. [95] used robot Pepper to investigate the distinctive design requirements for using LLMs in robots, that may be variable depending on the task and content. The user study included scenarios and 32 participants answering a questionnaire. Results indicated that LLM-based robots elevated expectations for sophisticated non-verbal cues.

Table 2 includes additional details about the selected literature (see Table 1) on LLM-based social robots, while Fig. 3 illustrates used social robots included in Table 2.

Although all studies referenced in this subsection focus on the integration of LLMs in social robots but not on the special education framework, all presented methodologies and results can be extended to provide broader implications for LLM-based social robots in special education. The integration of VR, human-robot interaction methodologies, used LLMs and selected social robots, presented in this subsection, reveal trends and guidelines that represent potential

Table 2 Integration of LLMs to social robots

Ref.	Social robot	LLM	Scope	Application	Limitations
[79]	Jubileo open-source simulated humanoid robot	GPT-4	Human-robot realistic interaction	English teaching VR game	Lacks sentiment analysis and dynamic corresponding facial expressions, not including automatic pronunciation error detection
[81]	Furhat robot	ChatGPT	Building a facilitation social robot	Speakers diarization for multiparty conversations by interacting with a robot	Not enough domain-specific knowledge to entirely understand real-world interaction, incorrect turn-taking probably due to the rules of English grammar
[82]	Pepper and NAO	GPT-3	LLMs-based human-robot interaction dialogue system	Conference participants-robot open dialogue on random topics to experience the possibilities and limitations of LLMs in live human-robot interaction systems	Based on cloud service Google Cloud speech-to-text and the NaoQi text-to-speech, more like a verbal approximation for text-based GTP-3
[83]	Furhat robot	GPT-3	LLM-based interactive robot presenter	Presentation of paintings in a museum to a set of participants	Retention tests revealed that objective learning outcomes from the interactive robot were limited
[84]	Furhat robot	GPT-3.5	Emotion generation in human-robot dialogue	Evaluated with 47 participants through a card sorting game specially designed to elicit emotions to evaluate the influence of emotional expressions of the robot on users	Delay in the robot's expression, server overload in cases, limited emotional categories, inability to generate long-term emotional responses
[85]	EVA robot	ChatGPT	Nonverbal communication	The robot recognizes the presence of ingredients on a plate and the LLM analyzes them to provide potential recipes	< not referenced>
[86]	Robot Mini	BERT, RoBERTa, Davinci for summarization, BERT, mT5, Davinci for question-answer	Personalized verbal human-robot interaction	Interaction with participants to evaluate the ability of the system to maintain an engaging personalized conversation	Similar profiles of participants, unable to update the static personal data of users
[87]	Robot Mini	GTP-3, T5, mT5, PEGASUS, and BERT2BERT	Natural conversational human-robot experience	Paraphrase generation and user-adapted semantic description approaches to allow free interaction or generate appropriate conversation topics	Paraphrasing may result in loss of meaning, interaction delays, limited computational power to run some LLMs
[88]	Furhat robot	GPT-3.5	Personalized companion robot	Interaction with 28 senior users to evaluate the human-robot interaction and identify primary obstacles	Hallucinations and obsolete information, and disengagement cues, resulting in confusion, frustration, and worry, robot interrupting the user, slow, superficial and repetitive
[89]	QTrobot	GPT-3	Personalized companion robot	Interaction with 12 senior users to evaluate the human-robot interaction and identify primary obstacles	Needs to improve flow of conversation, testing in real-world scenarios, integration of nonverbal cues
[90]	Tabletop Robot Haru	Llama-2-70B-chat	Dynamic and expressive conversations	12 human participants engaged in conversation for feedback and error analysis	Automatic speech recognition problems but LLM could recover, small class of LLM errors including hallucinations and repetitions
[91]	Furhat robot	CodeLlama, Llama2	Chat in English about Japanese cooking using a Japanese knowledge base	Case study with no participants	Viabilty of the approach to develop cooperative and multilingual social robot applications
[92]	Pepper	GPT-3	Robot autonomous responses to human verbal input	31 participants and three questionnaires to evaluate the experience	Longer interactions or more interactions were needed to decrease the variance
[93]	Furhat robot	GPT-3.5-turbo	Creation of virtual patients	Tested on 15 medical students through a Wilcoxon signed rank test to compare the Robot vs. a traditional approach	Minor problem of hallucinations

Table 2 (continued)

Ref.	Social robot	LLM	Scope	Application	Limitations
[94]	ARI robot	Alana v2	Multi-party conversations	Tested and improved through regular user tests where the robot acted as a receptionist in a hospital waiting room	No context-dependent gestures yet
[95]	Pepper	GPT-3.5	Human-robot interaction	A user study with 32 samples comparing an LLM-powered social robot against text- and voice-based agents through a mixed-factorial design with scenario tasks	LLM-powered robot was less preferred in one of the tasks, due to communication difficulties and the potential anxiety during collaboration

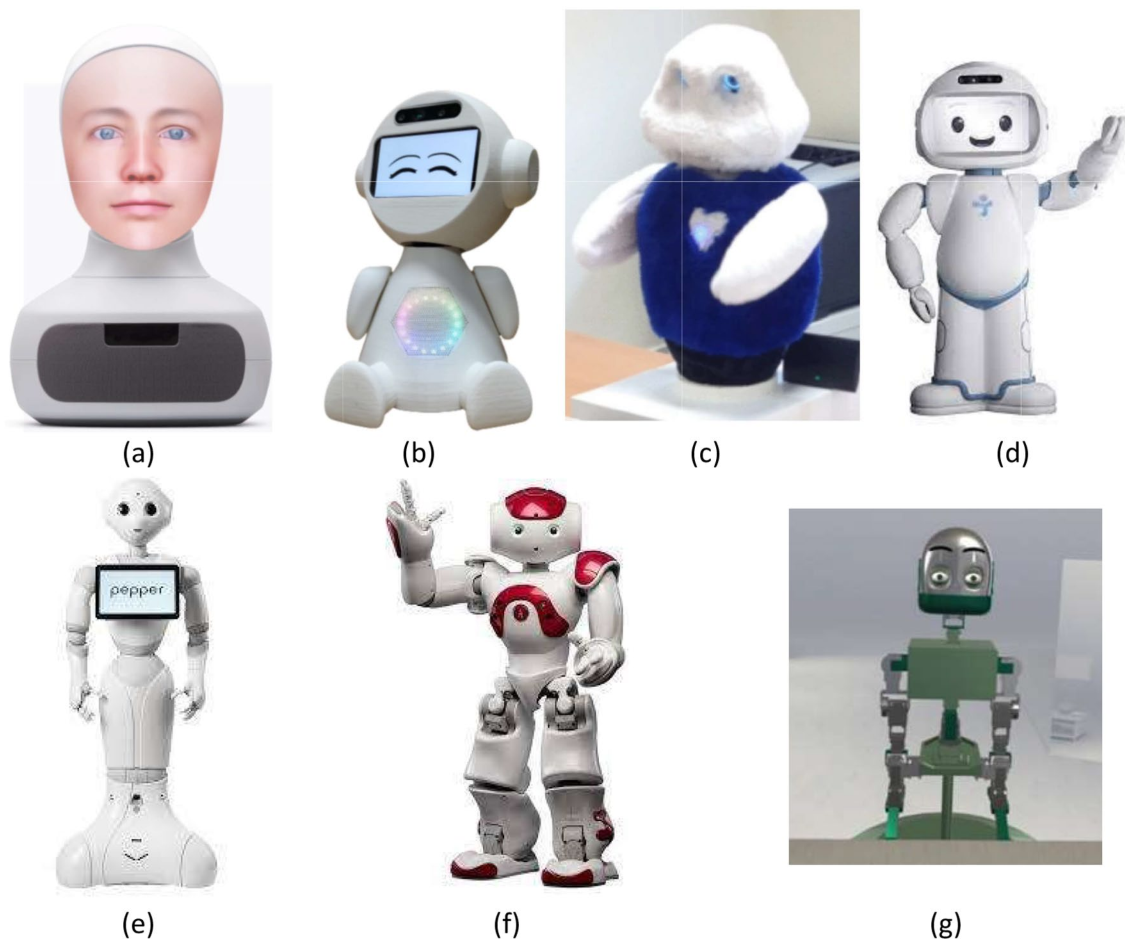


Fig. 3 Social robots empowered with LLMs according to the examined bibliography: (**a**) Furhat [102]; (**b**) EVA [103]; (**c**) Mini [104]; (**d**) QTrobot [105]; (**e**) Pepper [106]; (**f**) Nao [107]; (**g**) Jubileo [79]

and scalable tools to be extended and integrated into modern special education environments.

4.2 LLMs in special education (LLM& SE)

While LLMs are considered powerful transformative tools in the field of education [31], the subfield of special education has not yet fully embraced their capabilities. LLMs have the potential to offer assistance to educators by generating customized resources and planning lessons towards creating personalized education strategies to support students with special educational needs.

Applications of LLMs in special education involving testing and evaluation of specific LLM-based interventions to a group of students are not yet reported in the literature. However, LLMs have been proposed to provide support to students with disabilities, yet not in a systematic way (Table 3). Indicatively, the authors in [96] focused on evaluating LLM's engaging ability in empathetic, adaptable, and contextually suitable interactions during therapeutic interventions with hypothetical ASD impairment. The LLM

Table 3 Application of LLMs in special education

Ref.	Impairment group	Scope	Application	LLM	Limitations
[96]	High-functioning autistic adolescents	Simulated real-life therapeutic scenarios on hypothetical patients	Evaluation by experts of LLM's empathy, adaptability, communication, and engagement during therapy	Development of prompts to appropriately guide an LLM	Not consistent in complex emotional scenarios, not efficiently deep responses, varying levels of engagement
[97]	Children with Autism	LLMs for image classification	Classification of 200 images in autistic and non-autistic class (among other applications)	LLaVA 1.5 Large multimodal model	Lack of ability to process multiple images, produces hallucinations and misinformation
[98]	Attention Deficit Hyperactivity Disorder (ADHD)	AI writing workflow for reduced cognitive loads	A model of executive-cognitive capacity to assess how to manage the cognition of tasks and workloads, and support a design matrix for assistive tools and processes	GPT-3	No limitations were reported
[99]	Adults with dyslexia	Assisted email writing prototype	Evaluated in 19 adults with dyslexia	LaMDA	May not yet has sufficient accuracy to meet the needs of writers with dyslexia
[100]	Children with Central Auditory Processing Disorder (CAPD), and Visual Processing Disorder (VPD)	Assisting children with different learning styles such as visual learners or auditory learners	Focused on enabling persons with disabilities by tapping into the latest advances in AI, not tested	Multimodal Generative AI using Visual language models ViLT and GIT	No limitations were reported
[101]	People with motor and speech impairments	Social communication through eye gazing	Preliminary use case to boost social chat with gaze inputs to generates multiple sentences of conversation in real time based on the relationship of people in conversation	GPT-4	The communication phase, i.e., update and improve the process of content preferences and social closeness, needs more testing for its stability

was developed using a set of specially developed prompts to guide it through appropriate interaction with ASD adolescents. The performance evaluation of the LLM included empathy, skills to adapt and communicate, as well as engagement and abilities to launch a therapeutic-appropriate interaction, assessed by clinical psychologists and psychiatrists with varying levels of experience in ASD. Specific evaluation metrics were selected in line with the standards followed in psychological and autism therapy, forming an evaluation scorecard. The assessment was conducted by a panel of clinical phycologists and psychiatrist using the developed scorecard. The model could validate the emotions of patients; however, it was not always consistent. It could properly communicate, adapt, and respond in all cases and was revealed to be engaging even though its responses were not as deep as expected from a therapeutic session. Overall, the proposed approach revealed its potential as a supplementary tool in ASD therapy with adolescents, leaving room for further improvement. It should be noted here that the use of LLMs in special education should always be supplementary and used in the presence of experienced professionals who could adapt any technological tools to the specialized needs of each student's disability.

Islam et al. [97] explored the efficiency of LLMs to perform image classification Among others, a dataset (Autistic Children Facial Image Data Set) including images of faces of children with and without autism was used, concluding in 83% of correct classification accuracy after fine-tuning. Packer et al. [98] implemented a review article focusing on LLMs for people with Attention Deficit Hyperactivity Disorder (ADHD), Autism Spectrum Disorder (ASD), and other learning difficulties, exploring the cognitive load that was associated with complex writing tasks, and how the latter affected users with certain impairments. Goodman et al. [99] proposed an LLM-based inference to assist adults with dyslexia in writing emails. The interface was evaluated by 32 participants, regarding the frequency and duration of writing new emails and replies through a questionnaire. Results indicated the usefulness and consistency of the system (no numerical evaluation results were reported). In [100], the authors suggested an assisting tool based on LLMs for people with visual or auditory difficulties, able to dynamically adapt to strengths and abilities of the individual user. The authors mapped the challenges and proposed a design approach for their system; therefore, no evaluation performances were reported. Fang et al. [101] focused on people with motor and speech impairments to enhance their social communication skills with gaze inputs by using GPT-4. Separate user datasets were created, containing preferences and social closeness, and were used to generate tailored sentence suggestions for multi-turn conversation. A prototype test was conducted with three patients, reporting engagement and average conversation round around three topics more than 12, in 3 min.

Further utilization of LLMs in special education to empower students with special needs includes the following applications:

- LLMs can be combined with speech-to-text/ text-to-speech capabilities for people with *visual impairments* [108]. The latter possibility was recently launched in September 2023 by Meta and Ray-Ban, envisioning a supporting tool able to answer questions, summarize text, and read information to enhance the quality of life of the visually impaired. Future updates are expected to increase the efficient applicability of this integration further.
- Students with *hearing impairments* could benefit from LLMs in education by generating real-time texts, as well as,
- Students with *learning disabilities* could employ them to make the comprehension of complex texts easier [109].
- LLMs could be used combined with embedded devices to generate speech for students with *speech impairments*.

It should be also noted the novel introduction of Large Language and Vision Assistant models (LLVAs) that have also been incorporated in special education settings. Islam et al. [97] introduce LLVAs to detect children with autism from face images, underscoring the transformative potential of such models and their wide range of applications in real-world scenarios. For this reason, the latter application has been included in Table 3, although considered more as a supporting diagnostic tool.

As research on LLMs is ongoing, their application in special education is expected to evolve so that both students and teachers to benefit in practice.

4.3 LLM-based social robots in special education (LLM & SR & SE)

Based on all the above, it can be foreseen that the deployment of LLM-empowered social robots in special education holds significant potential. LLMs are anticipated to provide social robots with adequate AI awareness, empathy, and emotional adaptation to engage students and efficiently guide personalized interactions in special education. The conducted literature review revealed only two reported implementations of LLM-based social robots in special education, indicating that related research is still in its infancy (Table 4).

In their work, Lim et al. [110] enabled the social robot Pepper to understand American sign language towards enhancing nonverbal interaction for hearing impaired people. Even though the implementation of Lim et al. is not practically applied to special education settings, it is attributed to this field since it is the only related research to combine LLMs, social robots, and people with special needs. The authors developed a lightweight model for sign language recognition and produced context-aware gestures with Pepper using ChatGPT. The system was evaluated by empirical observations, aiming to highlight strengths and challenges. Results indicated the profound potential lying in human-robot interaction towards making technological tools, such as social robots, accessible for all. Mishra et al. [111] introduced an LLM pipeline with GPT-2 and BART towards generating text that the robot NAO could vocalize. The SOCIALIQA dataset was used for the generation

Table 4 Integration of LLMs to social robots for applications in special education

Ref.	Impairment group	Scope	Application	Social robot	LLM	Limitations
[110]	Hearing impairment	Nonverbal interaction	Social interaction with sign language recognition	Pepper	ChatGPT	Limited three-dimensional understanding in depth prediction
[111]	Autism	Teach perspective-taking in a therapy session using generated text	To generate texts that the robot vocalizes to use in a clinical setting in the presence of experts. The robot's role is of a stimulator, prompter, and reinforcer	NAO	GPT-2, BART	The system needs to be more versatile in scenarios to be taught

tasks. Context generation by using GPT-2 reported accuracy of 51%, indicating that the model generated data similar to the test set. BERTscore was used to evaluate question and option generation tasks, reporting high precision (up to 90%), recall (up to 91%) and F1-score (up to 90%). Statistical analysis of self-reports from domain experts was also considered. The robot was proposed to be used in a clinical setting in the presence of experts for autism interventions. The robot had multiple roles, e.g., of a stimulator, prompter, and reinforcer, out of which the stimulator role had autonomy in the text generation the robot used.

It should be noted, though, that this integration experiences limitations and challenges and brings forward ethical concerns regarding their development and application in special education settings, discussed in the following.

5 Discussion: challenges, limitations, and ethical considerations

In what follows, the challenges, limitations and considerations of applying LLM-based social robots in special education are identified, following a structure for this section as imposed by the proposed taxonomy (Table 1), aiming to provide answers to RQ2.

5.1 Challenges

Challenges refer to problems or difficulties arising during the process of integrating LLMs in social robots for special educational purposes, that need to be overcome. The latter may be robot-, LLMs-, or special education- related challenges.

5.1.1 Robot-related challenges

While NAO robot is one of the most commonly used social robots in special education, especially in applications with ASD students, early work on the integration of LLMs in social robots indicated preferable alternatives to NAO. Most research works included in Table 2 (37.5%) use the Furhat robot.

A powerful LLM alone is not adequate to develop an efficient human-robot interaction environment. Successful interaction with social robots also implies speech, voice, emotions, and body language. In human-robot interaction in special education, eye gaze is considered as a measure of engagement and joint attention. Moreover, since LLMs are engaging students in realistic conversations, social robots need to have a natural and adaptable appearance to match the conversation. However, NAO robot, along with several other popular social robots, conveys an emotionless facial expression. Furhat owns a back-projected face that permits facial expressions to be displayed, including movements of eyes and brows and head gestures. These capabilities enable the robot to communicate emotions, both perceived and conveyed. Emotional awareness of robots can be highly engaging since users are experiencing a more realistic and immersive interaction. The latter is essential in special education. Note that one of the reported limitations in Table 2 is the integration of nonverbal cues that are expected to improve the flow of conversation. Alternative robots that hold an animated faceplate are also Moxie, QTrobot, and EVA. The last two are also reported in Table 2, justifying the use of social robots that can display emotions as preferable when combined with LLMs.

Moxie [112] has been used for Mental, Behavioral, and Developmental Disorders as a rehabilitation tool for children. It is a socially interactive robot and a powerful therapeutic tool, as children are engaged in games and activities and interact with stories of Moxie. The overall interaction through emotion, speech, and expression can be a great asset in therapy intervention with children. It is worth noting that Moxie is connected to ChatGPT and further uses the cues of the interaction with the aid of GPT to be adjusted to the user's needs. Moxie, along with QTrobot, were also recommended as appropriate brain/mind-controlled robotics [113] for students with brain damage or neurological disorders who face difficulties in verbal communication. QTrobot recently integrated the latest LLM technologies, employing ChatGPT for language understanding and generation [105].

GPT models are considered to be the most notable and well-known LLM models, yet Bard, LLaMa 2-Chat, and other models are viable alternatives in the field of special education for social robots. Moreover, research should take into account other options for social robots, which already have been cooperating with ChatGPT, like LOONA [114], Cozmo [115], and Emo Robot [116]. LOONA has the potential for graphic programming and its main module, GPT Wonderland, encompasses a lot of activities for children. Ameca is thought to be the most advanced humanoid robot at the moment [117], while OpenAI and 1X have partnered to produce a robot that will incorporate GPT-5 [118].

While one aspect of the LLMs integration into robots is the appropriate robot's design, another is the robot's voice. Mini robot is designed to interact in Spanish, therefore LLMs integration in Mini needs to also include a module to translate prompts from/to Spanish. Multilingual models need to be adopted in such cases, to also include alternative languages at their corpus. However, as reported in the examined literature, the translation of one language to another and back, may result in paraphrasing and distort the meanings.

One additional constraint for LLMs to be integrated into social robots is related to the computational power required for LLMs to run in real-time, either locally or in the cloud. Most social robots have limited resources, while several other modules are simultaneously running on the robots along with LLMs, e.g., for vision tasks and motor control, leading to either processing delays or server overload in some cases. The latter is the reason for the reported limitations in efficient human-robot interaction, causing delays in the robot's dynamic facial and verbal expressions, translated as difficulties from LLMs to provide adequate empathy and adaptability. The achievement of real-time interaction can be quite challenging when using LLMs; natural language processing in real-time while retaining responsive actions requires both efficient algorithms and hardware. Finally, LLMs are mainly trained on text data and may not inherently comprehend the physical context in which social robots operate. Integrating sensory input and contextual awareness remains a challenge.

Robot selection in all cases depends on the application and the target group. Not all robots display the same capabilities, while others are specially designed to support certain age groups (as Mini for the elderly), or certain impairments (as PARO for dementia, or NAO and Milo for autism). Nowadays, several types of social robots are available in the market, each with different features and capabilities to interact with humans. The selection of the appropriate robot must rely on its suitability for each target group, as well as its LLM integration capabilities in terms of technical characteristics. Therefore, the LLM-based social robot selection must be performed by considering technical and pedagogical criteria, specific to each educational scenario and student target group. Selecting the appropriate type of social robot can be handled as a multi-criteria decision-making (MCDM) problem, considering technical factors (e.g., autonomy), capacity (e.g., memory), economic factors (e.g., costs), interaction capabilities (e.g., reported improvement), social factors (e.g., acceptance), and more, that need be equally weighted under the prism of user impairment characteristics (e.g., autism characteristics) towards indicating the most appropriate robot for each use case.

An additional robot-related challenge concerns the integration of multiple models on the robot, and their synchronization. While LLMs are used for communication, other models are simultaneously run on the robot for behavioral and emotional analysis. Their synchronization can be achieved through modular architectures of the robots' software. Modular architectures allow each module to operate separately its specific function, while it communicates with other modules through well-defined interfaces. Real-time processing is required to analyze facial expressions, vocal tones, posture, etc., and determine the overall status of the user. The robot, thus, collects and analyzes data from its sensors continuously and outputs users' emotions, behaviors, and more. A central control system gathers the outputs of all modules and uses them as inputs to drive the LLMs context, and subsequently the robot actions. Finaly, feedback loops are used to adjust the robots' language and gestures, so as to match the LLMs response. The latter process is challenging due to (1) the complexity of integration resulting from the different technical demands of each model, (2) the real time processing requirement so as to handle computational loads and provide fluidity in the human-robot interaction, (3) the difficulties involved in contextual understanding, (4) and the inherent hardware limitations in movements and facial expressions of robots, not always aligning with the software output.

5.1.2 LLMs-related challenges

In some cases, LLMs can be harmfully biased or overfitted, leading to hallucinations, repetitions and superficial conversations, as seen in the reported limitations in Table 2. Additionally, LLMs may generate content that, although not clearly censored, could be considered unsuitable for special education applications. The latter results in frustration and confusion, provoking negative emotions in the users. Since the emotional stability of children is a key component to promoting special education initiatives, the LLMs need to be as reliable as possible. For this reason, it should be investigated whether the use of emotion recognition software [119] in combination with appropriate prompt engineering, can ensure proper LLM alignment, i.e., ensuring that the LLM can act safely and as expected.

In general, the understanding and interpretation of LLMs' outputs could be challenging. Transparent decision-making is crucial in social robotics, especially when dealing with sensitive or critical tasks such as those related to special education. The users' acceptance and trust, referring to both students and special educators, are clearly affected by the performance of LLMs on the robots. Unexpected responses from LLMs can impact user acceptance, emphasizing the need for careful design and user experience considerations.

At this point, it is worth noting that the literature lacks real-world scenarios' applications. Evaluation of LLM-based social robots in special education has not been reported yet. Therefore, the fidelity of interventions with LLM-based social robots is not measured, restricting the ability to study the practical outcomes of their use.

5.1.3 Special education-related challenges

LLMs designated for special education need to be trained with specific data. LLMs are developed to understand and generate human language in the typical way. However, communication with students with special educational needs is far from typical. Each impairment group has different ways to express and communicate and has different needs, e.g., ASD children cannot engage and lack verbal and nonverbal communication skills, children with dyslexia have difficulty expressing themselves through speech but do not have learning difficulties or lack of communicational skills, children with hearing impairments cannot process linguistic information but can process nonverbal cues, etc. It is, therefore, evident that since each impairment group has different needs, distinct LLMs need to be developed specifically for each group and integrated with various social robots tailored to those groups [120]. The latter requires domain-specific knowledge that needs to be integrated into the models in terms of verbal/nonverbal cues and appropriate educational scenarios for each impairment. Fine-tuning of LLMs for specific social groups requires labeled datasets and expertise. Acquisition and annotation of diverse datasets for multiple social scenarios is expected to be challenging.

Moreover, there is a need to maintain an equilibrium between standard responses and personalized educational interventions. The LLM-based robot needs to maintain both general coherent and personalized dialogues to resemble a human therapist. The challenge in special education is not only to deliver an LLM-based robot that simply understands and gives feedback but also to tailor it to specific individual needs.

5.2 Limitations

Limitations refer to constraints or boundaries restricting the process of integrating LLMs in social robots for special educational purposes, i.e., the restrictions within which this integration must operate. Limitations can also be robot-, LLM-, and special education- related. Limitations are closely related to the aforementioned challenges, and they only differ in the sense that challenges can be overcome whereas limitations set boundaries that cannot be overcome. Thus, limitations are considered challenges that cannot be overcome.

By thoroughly examining the challenges of the integration of LLMs in social robots for special education purposes, inherent limitations have been identified. In what follows, these main limitations are briefly mentioned.

5.2.1 Robot-related limitations

Robot-related limitations are related to the physical and social interaction limits provided by machines such as robots. Even by using the most advanced LLM, social robots could not fully replicate the wide range of human social interactions and emotions, communicated through facial expressions and complex gestures, that are essential in special education educational interventions, especially for children with social and emotional impairments.

5.2.2 LLM-related limitations

Advanced LLMs can easily stimulate a conversation, yet there is an inherent difficulty in the true understanding of the needs of students with special needs. The latter poses limitations in the interactions, due to the lack of a deep comprehension of a student's emotions and needs.

Moreover, each students' needs differ significantly even among the same impairments; there are several scales to evaluate autism, since each case has a different severity and requires different therapeutic approaches. Thus, LLMs may appear empathetic, yet they fall short in fully supporting different students with complex variable emotional and social needs.

LLMs require large datasets to train upon each special impairment case in special education so as to function effectively. Data privacy and related ethics, though, pose limitations in acquiring such sensitive personal data, limiting the capabilities of personalized feedback of the robot. Even in case of accessing inclusive data, the fairness of an LLM-based robot could be limited from biased training data.

5.2.3 Special education-related limitations

LLMs are not endorsed to make therapeutic or educational decisions, therefore they cannot fully replace a human therapist or teacher in special educational settings.

Another limitation is related to accessibility and costs of acquiring and maintaining an LLM-based social robot. Special education institutions may struggle to have access to such technologies, especially in underfunded or rural regions, e.g., rugged areas, remote mountainous villages or islands.

The identified limitations are summarized along with their implications in Table 5.

5.3 Ethical considerations

As already mentioned, LLMs can generate content that might be biased or inappropriate. In special education, the latter could lead to unfair treatment or misinterpretation of the needs and emotions of students with disabilities. All LLMs-generated tasks need to obey basic ethical and educational standards, ensuring that the models are trained on unbiased data and that they would not generate discriminative or unfair results for persons with special needs. Biased AI models can influence critical decision-making. However, it is challenging to fully comprehend how AI models make those decisions, especially in complex LLM models.

Table 5 Limitations of integrating LLMs in social robots for special educational purposes

Category	Description of limitation	Implications
Robot-Related Limitations	**Physical Interaction Limits**: Social robots cannot fully replicate the complexity of human expressions and gestures, which are crucial in special education.	Limits the effectiveness of emotional engagement and interaction, particularly for students with social impairments.
	Hardware Constraints: Limited processing power in social robots restricts the real-time processing capabilities required for dynamic LLM interactions.	Delays in responses can lead to less natural interactions, reducing the robots' effectiveness as empathetic companions.
LLM-Related Limitations	**Lack of True Emotional Understanding**: LLMs lack genuine emotional comprehension, resulting in responses that may not fully resonate with students' needs.	Can impact the quality of interactions, particularly for students requiring sensitive, personalized engagement.
	Data Privacy and Ethical Concerns: Access to large, diverse datasets for training LLMs in special education settings is restricted by privacy concerns.	Limits personalization and adaptability of responses; raises concerns about sensitive data protection.
	Inconsistent and Biased Outputs: LLMs may produce biased or inappropriate content due to training data limitations, especially in sensitive contexts.	May lead to confusion or frustration among students; requires educator oversight to prevent potentially harmful interactions.
Special Education-Related Limitations	**Limited Accessibility in Underfunded Regions**: High costs and maintenance requirements limit access to LLM-based social robots in certain areas.	Reduces the potential for equitable access to this technology in special education, especially in rural or remote locations.
	Inadequate Fit for Diverse Needs: Students with special needs vary greatly in their requirements, which makes it difficult for LLMs to cater to all needs effectively.	LLMs cannot fully replace human educators or therapists, particularly in complex cases requiring nuanced understanding.
General Limitations	**Complex System Integration**: Real-time synchronization of LLMs with emotional and behavioral analysis models is technically challenging.	Reduces system reliability and response fluidity, impacting the robot's ability to deliver seamless, adaptive interactions.

Explainable and responsible development of AI needs to follow specific guidelines and regulatory standards, such as the IEEE Ethically Aligned Design [121] or the AI Ethics Guidelines from the European Commission [122]. Human inspection of the LLMs outputs is required to identify any possible biases and interfere accordingly. Therefore, the role of special educators is preserved at all times; LLM-based social robots are assistive tools and should always be used under their supervision and responsibility.

The use of such technological tools involves the acquisition and processing of students' sensitive personal data, raising questions about their management. As LLMs constantly evolve, the development of ethical guidelines regarding data privacy and security policies is necessary for protecting data from authorized access and providing students and their parents the necessary transparency on the way data is handled before consenting to participate in any intervention. Informed consent from students and their guardians is necessary, as to be fully aware of how data is used, stored and protected. Data security measures through encryption techniques, secure data storage and regular security inspections, could safeguard sensitive information from unauthorized access. Since the models designated for special education need to be trained on private data for personalized treatments, additional ethical concerns arise in case the developed models are later available publicly. Data protection regulations such as General Data Protection Regulation (GDPR) [123] and Consumer Online Privacy Rights Act (COPRA) [124], need to be obeyed, towards ensuring the ethical use of LLMs in education through the provision of guidelines for handling and protecting sensitive student data.

The phycological impact on students due to these ethical implications is significant [125]. Lack of trust and safety can lead to anxiety and refusal to participate in LLM-based robotic interventions. Perceived bias can affect self-esteem and create feelings of frustration and marginalization. Decision making from a non-explainable AI system could make students feel powerless and out of control over the processes. Moreover, the generated stress over these new technologies can overall affect their mental health [126, 127].

Considering ethical and safety restrictions, as mentioned above, a safe and stable environment should be created so that children with special needs can be effectively treated. Additionally, research has uncovered the scientific community's limited interest in children's mental safety when interacting with AI technologies [128]. This natural imitation of a human-like voice can create emotional bonds with children. Nevertheless, issues of toxic dependencies should be addressed. The prime objective is the social integration of children with special educational needs in society, not any form of bonding, which could result in self-isolation.

6 A framework for LLM-based social robots' integration in special education

Addressing the aforementioned challenges involves interdisciplinary efforts involving expertise in robotics, natural language processing, special education, and ethics, aiming to contribute towards the successful integration of LLMs for social robots in special education.

To this end, a framework of LLM-based social robots in special education needs to be formulated to ensure that their conscious use would only be beneficial. Implementing LLMs via social robots in the field of special education is a demand which is still under-researched. The mainstream design approach is through Cloud (GPT-J) since cloud services can ensure the privacy of data [129]. Whether children have SEN or not, when they are exposed to AI, they have to be protected due to the vulnerability of their age. For this reason, recently, children's fundamental rights in human-robot Interaction have been developed based on UNICEF's AI Policy Guidance [128].

Every impairment in children with special educational needs may vary, causing diverse symptoms; moreover, multiple disorders may coexist in the same person. Therefore, guidelines should be formed and adopted by LLMs, addressing limitations, learning strategies, emotional and psychological approaches, and more. LLMs are the novel key element of communication within a social robot. The robot should take various actions, like initiating a conversation, maintaining it, evaluating, explaining, and more, within the learning process. This learning cycle of steps must be properly adapted, planned, and executed. Before applying LLMs in the field of special education, it is of high importance to enumerate all their potential functions, which are particularly useful for children with special needs. Such students may often have limited or zero verbal communication or may undergo outbursts of negative emotions, e.g., anger or disappointment. The benefits of the integration of LLMs in social robots are clearly evident, yet the type of intervention a system should adopt in times of an ongoing crisis is still a matter of debate.

Based on all the above, it is evident that integrating an LLM-based social robot into special education requires a thoughtful framework to address the unique needs and challenges of students with diverse learning requirements. In what follows, aiming to provide answers to RQ3, a framework is proposed, involving 20 subsequent steps for three different roles:

6.1 Educators

1. **Needs assessment and goal setting**: First, a thorough assessment needs to be conducted involving special education experts to identify specific learning goals, challenges, and objectives for students with different abilities and needs. Such an assessment includes a set of specialists, such as special education teachers, psychologists, speech therapists, etc., provide their valuable insights regarding the special needs of each individual student.
2. **Customization and personalization**: Design the LLM integration to allow for customization and personalization. Tailor the learning experience based on individualized special education plans, learning styles, and progress tracking for each student. Therefore, different goals would be set for each student, tailored to their abilities and challenges.
3. **Align with curriculum standards**: Ensure that the LLM aligns with special education curriculum standards, facilitating seamless integration into existing special education programs and lesson plans.
4. **Environmental considerations**: Educators need to access the setup of the classroom to make sure that it is appropriate to accommodate sensory needs of both students and robots and minimizes their distractions.

6.2 Developers

5. **Social robot selection**: Based on the different abilities and needs of each student, as well as the different learning goals, the appropriate social robot is selected, so as to possess the requirements for delivering the designed special education plan.
6. **Accessible user interface**: Develop an accessible and user-friendly interface that accommodates various assistive technologies, ensuring that students with diverse abilities can interact effectively with the LLM-based social robot and the special teachers can easily operate.
7. **Multimodal interaction**: Implement multimodal interaction capabilities, incorporating visual, auditory, and tactile elements to support students with different sensory preferences and needs for both verbal and nonverbal interaction.
8. **Adaptive learning approaches**: Integrate adaptive learning technologies that can adjust the content, pace, and difficulty levels based on the progress of individual students.

9. **Speech and gesture recognition and feedback**: Incorporate robust speech and gesture recognition capabilities to facilitate communication for students with language and hearing or other impairments. Provide constructive and immediate feedback to support learning.
10. **Emotional intelligence**: Embed emotional intelligence features into the LLM to recognize and respond appropriately to the emotional states of students, fostering a supportive and empathetic learning environment.
11. **Data privacy and security**: Implement robust data privacy and security measures to protect students' sensitive information. Comply with relevant regulations and guidelines to ensure the ethical use of special student data.
12. **Regular assessment and progress monitoring**: Integrate assessment tools within the LLM and by using embedded sensors on the social robot to monitor students' progress. Provide real-time feedback to educators to support ongoing improvement.
13. **Assistive technology integration**: Ensure compatibility of the system and ability to integrate with other assistive technologies, such as VR, to address the diverse needs of students with disabilities.
14. **Continuous improvement and updates**: Establish a framework for continuous improvement, regularly updating both the LLMs' and the robots' capabilities based on feedback, emerging research, and advancements in technology.

6.3 Stakeholders

15. **Expertise collaboration and support**: Provide resources and training for educators to integrate the LLM-based social robot into their teaching practices effectively. Encourage collaboration between teachers, special education professionals, and technology experts.
16. **Access to technology.** Ensure that all special educational settings are fully equipped with the necessary hardware and software to support such initiatives.
17. **User training and resources**: Develop user-friendly training materials and resources, e.g., manuals and tutorials, for both educators and students. Provide ongoing support to ensure effective utilization of the LLM-based social robot in special education settings.
18. **Professional development**: Create and offer continuous professional development opportunities to inspire and motivate educators, so as to be kept updated on the latest advancements and best practices.
19. **Parents' involvement and communication**: Introduce the LLM-based social robot to parents and promote communication between them to gain their trust. Provide insights into a child's progress, learning achievements, and areas for improvement to encourage collaboration between home and school.
20. **Dissemination of research results**: Share relevant studies and research findings regarding the effectiveness of LLM-based social robots in education among educators and parents to strengthen their trust on such initiatives.

The proposed framework emphasizes the importance of collaboration, customization, and ongoing support to create a safe, positive, and inclusive learning environment for students in special education, and suggests actionable steps that could be followed by educators, developers and stakeholders. Yet, those guidelines mean to be general allowing them to be adaptable and applicable across various contexts and situations, providing a broader structure that can be tailored to specific needs, and accommodate different interpretations and implementations.

7 Conclusions

As technology is evolving, its involvement in special education aims to offer transformative solutions. On the one hand, the exploitation of the impact of LLMs in the realm of special education yielded promising perspectives. On the other hand, the use of social robots in special education has proven their effectiveness. Combining both technologies by delivering LLM-based social robots to support special education is expected to revolutionize the special education sector. This approach may offer engaging, adaptable, and empathetic human-robot interactions and promote therapeutic, social, or emotional means of rehabilitation, a claim that needs further scientific research.

This work aims to identify the status and the potential of LLMs for social robots in special education, underlining all related challenges, limitations, opportunities, and ethical considerations, aiming to provide insight and generate guidelines for the use of LLM-based social robots towards their efficient integration into special education practices. The foreseen potential of LLM-based social robots has a long way to go until they are practically implemented and evaluated in educational settings. Our findings revealed that there is a lack of practical in-field implementations of LLM-based social robot in special education, while the development of related applications is in its early stages. Therefore, there is a need for constant research in the AI field, focusing on delivering student-centric LLM models that can be integrated into social robots to better meet the complex needs of special education students. Related challenges have been identified, including personalization issues, real-time responsiveness and emotional understanding, alinement and hallucinations, along with limitations related to the physical and social interaction limits of robots, related costs and accessibility of such supporting technologies that impose constraints on the effectiveness of use of LLM-based social robots in special education, as well as ethical considerations, related to data privacy and inherent biases. The latter challenges represent problems that could be addressed with further research and innovation, while limitations and ethical considerations impose constraints on the effectiveness and safety of LLM-based robots in special education. These limitations and considerations require attention and strict human oversight, so as to mitigate all related risks affecting their safe and efficient application.

This work concludes by delivering a framework for the integration of LLM-based social robots in special education, as ethical considerations, best practices, effective AI-based learning strategies, and more factors need to be considered in their design process. The proposed framework aims to address the unique needs and challenges of students with diverse learning requirements, involving 20 subsequent steps for three different roles, i.e., of educators, developers and stakeholders, and deliver the first reported guidelines for the integration of LLM-based social robots in special education settings. The proposed framework is considered a valuable contribution towards the smooth integration of LLMs-based social robots in special education.

Future work includes in-depth pilot studies in special educational settings to assess the practical application of an LLM-based social robot by following the proposed framework. Addressing reported challenges will also be on focus, towards enhanced personalization aiming to the development of a safe and effective educational tool intended for special education.

Funding Open access funding provided by HEAL-Link Greece.

Declarations

Competing interests The authors report there are no competing interests to declare.

References

1. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are Few-Shot Learners. Adv. Neural Inf. Process. Syst. **33**, 1877–1909 (2020)
2. Ackerman, R., Balyan, R.: Automatic Multilingual Question Generation for Health Data Using LLMs. Presented at the (2024)
3. Zhang, T., Ladhak, F., Durmus, E., Liang, P., McKeown, K., Hashimoto, T.B.: Benchmarking Large Language Models for News Summarization. arXiv Prepr. arXiv.2301, (2023)
4. Nicula, B., Dascalu, M., Arner, T., Balyan, R., McNamara, D.S.: Automated Assessment of comprehension strategies from Self-explanations using LLMs. Information. **14**, 567 (2023). https://doi.org/10.3390/info14100567
5. Abburi, H., Suesserman, M., Pudota, N., Veeramani, B., Bowen, E., Bhattacharya, S.: Generative AI text classification using ensemble LLM approaches. arXiv Prepr arXiv:2309.07755, (2023)
6. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida,D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji,S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., Bello, I., Berdine,J., Bernadett-Shapiro, G., Berner, C., Bogdonoff, L., Boiko, O., Boyd, M., Brakman,A.-L., Brockman, G., Brooks, T., Brundage, M., Button, K., Cai, T., Campbell, R.,Cann, A., Carey, B., Carlson, C., Carmichael, R., Chan, B., Chang, C., Chantzis, F.,Chen, D., Chen, S., Chen, R., Chen, J., Chen, M., Chess, B., Cho, C., Chu, C., Chung,H.W., Cummings, D., Currier, J., Dai, Y., Decareaux, C., Degry, T., Deutsch, N., Deville,D., Dhar, A., Dohan, D., Dowling, S., Dunning, S., Ecoffet, A., Eleti, A., Eloundou,T., Farhi, D., Fedus, L., Felix, N., Fishman, S.P., Forte, J., Fulford, I., Gao, L.,Georges, E., Gibson, C., Goel, V., Gogineni, T., Goh, G., Gontijo-Lopes, R., Gordon,J., Grafstein, M., Gray, S., Greene, R., Gross, J., Gu, S.S., Guo, Y., Hallacy, C.,Han, J., Harris, J., He, Y., Heaton, M., Heidecke, J., Hesse, C., Hickey, A., Hickey,W., Hoeschele, P., Houghton, B., Hsu, K., Hu, S., Hu, X., Huizinga, J., Jain, S.S.,Jain, S.S., Jang, J., Jiang, A., Jiang, R., Jin, H., Jin, D., Jomoto, S., Jonn, B.,Jun, H., Kaftan, T., Kaiser, Ł., Kamali, A., Kanitscheider, I., Keskar, N.S., Khan,T., Kilpatrick, L., Kim, J.W., Kim, C., Kim, Y., Kirchner, H., Kiros, J., Knight,M., Kokotajlo, D., Kondraciuk, Ł., Kondrich, A., Konstantinidis, A., Kosic, K., Krueger,G., Kuo, V., Lampe, M., Lan, I., Lee, T., Leike, J., Leung, J., Levy, D., Li, C.M.,Lim, R., Lin, M., Lin, S., Litwin, M., Lopez, T., Lowe, R., Lue, P., Makanju, A.,Malfacini, K., Manning, S., Markov, T., Markovski, Y., Martin, B., Mayer, K., Mayne,A., McGrew, B., McKinney, S.M., McLeavey, C., McMillan, P., McNeil, J., Medina, D.,Mehta, A., Menick, J., Metz, L., Mishchenko, A., Mishkin, P., Monaco, V., Morikawa,E., Mossing, D., Mu, T., Murati, M., Murk, O., Mély, D., Nair, A., Nakano, R., Nayak,R., Neelakantan, A., Ngo, R., Noh, H., Ouyang, L., O'Keefe, C., Pachocki, J., Paino,A., Palermo, J., Pantuliano, A., Parascandolo, G., Parish, J., Parparita, E., Passos,A., Pavlov, M., Peng, A., Perelman, A., Peres, F. de A.B., Petrov, M., Pinto, H.P.de O., Michael, Pokorny, Pokrass, M., Pong, V., Powell, T., Power, A., Power, B.,Proehl, E., Puri, R., Radford, A., Rae, J., Ramesh, A., Raymond, C., Real, F., Rimbach,K., Ross, C., Rotsted, B., Roussez, H., Ryder, N., Saltarelli, M., Sanders, T., Santurkar,S., Sastry, G., Schmidt, H., Schnurr, D., Schulman, J., Selsam, D., Sheppard, K.,Sherbakov, T., Shieh, J., Shoker, S., Shyam, P., Sidor, S., Sigler, E., Simens, M.,Sitkin, J., Slama, K., Sohl, I., Sokolowsky,

B., Song, Y., Staudacher, N., Such, F.P.,Summers, N., Sutskever, I., Tang, J., Tezak, N., Thompson, M., Tillet, P., Tootoonchian,A., Tseng, E., Tuggle, P., Turley, N., Tworek, J., Uribe, J.F.C., Vallone, A., Vijayvergiya,A., Voss, C., Wainwright, C., Wang, J.J., Wang, A., Wang, B., Ward, J., Wei, J., Weinmann,C., Welihinda, A., Welinder, P., Weng, J., Weng, L., Wiethoff, M., Willner, D., Winter,C., Wolrich, S., Wong, H., Workman, L., Wu, S., Wu, J., Wu, M., Xiao, K., Xu, T.,Yoo, S., Yu, K., Yuan, Q., Zaremba, W., Zellers, R., Zhang, C., Zhang, M., Zhao, S.,Zheng, T., Zhuang, J., Zhuk, W., Zoph, B., OpenAI,:, Achiam, J., Adler, S., Agarwal,S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S.,Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao,H., Bavarian, M., Belgum, J., Bello, I., Berdine, J., Bernadett-Shapiro, G., Berner,C., Bogdonoff, L., Boiko, O., Boyd, M., Brakman, A.-L., Brockman, G., Brooks, T.,Brundage, M., Button, K., Cai, T., Campbell, R., Cann, A., Carey, B., Carlson, C.,Carmichael, R., Chan, B., Chang, C., Chantzis, F., Chen, D., Chen, S., Chen, R., Chen,J., Chen, M., Chess, B., Cho, C., Chu, C., Chung, H.W., Cummings, D., Currier, J.,Dai, Y., Decareaux, C., Degry, T., Deutsch, N., Deville, D., Dhar, A., Dohan, D.,Dowling, S., Dunning, S., Ecoffet, A., Eleti, A., Eloundou, T., Farhi, D., Fedus,L., Felix, N., Fishman, S.P., Forte, J., Fulford, I., Gao, L., Georges, E., Gibson,C., Goel, V., Gogineni, T., Goh, G., Gontijo-Lopes, R., Gordon, J., Grafstein, M.,Gray, S., Greene, R., Gross, J., Gu, S.S., Guo, Y., Hallacy, C., Han, J., Harris,J., He, Y., Heaton, M., Heidecke, J., Hesse, C., Hickey, A., Hickey, W., Hoeschele,P., Houghton, B., Hsu, K., Hu, S., Hu, X., Huizinga, J., Jain, S.S., Jain, S.S., Jang,J., Jiang, A., Jiang, R., Jin, H., Jin, D., Jomoto, S., Jonn, B., Jun, H., Kaftan,T., Kaiser, Ł., Kamali, A., Kanitscheider, I., Keskar, N.S., Khan, T., Kilpatrick,L., Kim, J.W., Kim, C., Kim, Y., Kirchner, H., Kiros, J., Knight, M., Kokotajlo, D.,Kondraciuk, Ł., Kondrich, A., Konstantinidis, A., Kosic, K., Krueger, G., Kuo, V.,Lampe, M., Lan, I., Lee, T., Leike, J., Leung, J., Levy, D., Li, C.M., Lim, R., Lin,M., Lin, S., Litwin, M., Lopez, T., Lowe, R., Lue, P., Makanju, A., Malfacini, K.,Manning, S., Markov, T., Markovski, Y., Martin, B., Mayer, K., Mayne, A., McGrew,B., McKinney, S.M., McLeavey, C., McMillan, P., McNeil, J., Medina, D., Mehta, A.,Menick, J., Metz, L., Mishchenko, A., Mishkin, P., Monaco, V., Morikawa, E., Mossing,D., Mu, T., Murati, M., Murk, O., Mély, D., Nair, A., Nakano, R., Nayak, R., Neelakantan,A., Ngo, R., Noh, H., Ouyang, L., O'Keefe, C., Pachocki, J., Paino, A., Palermo, J.,Pantuliano, A., Parascandolo, G., Parish, J., Parparita, E., Passos, A., Pavlov, M.,Peng, A., Perelman, A., Peres, F. de A.B., Petrov, M., Pinto, H.P. de O., Michael,Pokorny, Pokrass, M., Pong, V., Powell, T., Power, A., Power, B., Proehl, E., Puri,R., Radford, A., Rae, J., Ramesh, A., Raymond, C., Real, F., Rimbach, K., Ross, C.,Rotsted, B., Roussez, H., Ryder, N., Saltarelli, M., Sanders, T., Santurkar, S., Sastry,G., Schmidt, H., Schnurr, D., Schulman, J., Selsam, D., Sheppard, K., Sherbakov, T.,Shieh, J., Shoker, S., Shyam, P., Sidor, S., Sigler, E., Simens, M., Sitkin, J., Slama,K., Sohl, I., Sokolowsky, B., Song, Y., Staudacher, N., Such, F.P., Summers, N., Sutskever,I., Tang, J., Tezak, N., Thompson, M., Tillet, P., Tootoonchian, A., Tseng, E., Tuggle,P., Turley, N., Tworek, J., Uribe, J.F.C., Vallone, A., Vijayvergiya, A., Voss, C.,Wainwright, C., Wang, J.J., Wang, A., Wang, B., Ward, J., Wei, J., Weinmann, C., Welihinda,A., Welinder, P., Weng, J., Weng, L., Wiethoff, M., Willner, D., Winter, C., Wolrich,S., Wong, H., Workman, L., Wu, S., Wu, J., Wu, M., Xiao, K., Xu, T., Yoo, S., Yu,K., Yuan, Q., Zaremba, W., Zellers, R., Zhang, C., Zhang, M., Zhao, S., Zheng, T.,Zhuang, J., Zhuk, W., Zoph, B.: GPT-4 Technical Report. arXiv Prepr. arXiv:2303, (2023)
7. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., Lample, G.: LLaMA: Open and efficient Foundation Language models. arXiv Prepr arXiv 2302. (2023). https://doi.org/10.48550/arXiv.2302.13971
8. Stade, E., Wiltsey Stirman, S., Ungar, L.H., Boland, C.L., Schwartz, H.A., Yaden, D.B., Sedoc, J., DeRubeis, R., Willer, R., Eichstaedt, C.: Large language models could change the future of behavioral healthcare: A proposal for responsible development and evaluation. PsyArXiv Prepr. (2023). https://doi.org/10.31234/osf.io/cuzvr
9. Arcan, M., Niland, P.-D., Delahunty, F.: An Assessment on Comprehending Mental Health through Large Language Models. arXiv Prepr. arXiv:2401, (2024)
10. Mao, C., Xu, J., Rasmussen, L., Li, Y., Adekkanattu, P., Pacheco, J., Bonakdarpour, B., Vassar, R., Shen, L., Jiang, G., Wang, F., Pathak, J., Luo, Y.: AD-BERT: Using pre-trained language model to predict the progression from mild cognitive impairment to Alzheimer's disease. J. Biomed. Inf. **144**, 104442 (2023). https://doi.org/10.1016/j.jbi.2023.104442
11. Xu, L., Mueller, K.D., Liss, J., Berisha, V.: Decorrelating Language Model Embeddings for Speech-Based Prediction of Cognitive Impairment. In: ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023)
12. Mukherjee, P., Sadhukhan, S.G.R., Godse, S., Chakraborty, M.: Detection of Autism Spectrum disorder (ASD) from Natural Language text using BERT and ChatGPT models. Int. J. Adv. Comput. Sci. Appl. **14** (2023). https://doi.org/10.14569/IJACSA.2023.0141041
13. Liu, N., Wang, Z., Baraniuk, R.G., Lan, A.: GPT-based Open-Ended knowledge tracing. arXiv Prepr. arXiv:2203.03716, (2022)
14. Li, C., Xing, W., Leite, W.: Building socially responsible conversational agents using big data to support online learning: A case with Algebra Nation. Br. J. Educ. Technol. **53**, 776–803 (2022). https://doi.org/10.1111/bjet.13227
15. Sarsa, S., Denny, P., Hellas, A., Leinonen, J.: Automatic Generation of Programming Exercises and Code Explanations Using Large Language Models. In: Proceedings of the 2022 ACM Conference on International Computing Education Research - Volume 1. pp. 27–43. ACM, New York, NY, USA (2022)
16. Wang, Z., Lan, A., Baraniuk, R.: Math Word Problem Generation with Mathematical Consistency and Problem Context Constraints. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. pp. 5986–5999. Association for Computational Linguistics, Stroudsburg, PA, USA (2021)
17. Scarlatos, A., Lan, A.: Tree-Based Representation and Generation of Natural and Mathematical Language. arXiv Prepr. arXiv:2302, (2023)
18. Cooper, G.: Examining Science Education in ChatGPT: An exploratory study of Generative Artificial Intelligence. J. Sci. Educ. Technol. **32**, 444–452 (2023). https://doi.org/10.1007/s10956-023-10039-y
19. Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., Liu, T.-Y.: BioGPT: Generative pre-trained transformer for biomedical text generation and mining. Brief. Bioinform. **23** (2022). https://doi.org/10.1093/bib/bbac409
20. Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., Tanaka, F.: Social robots for education: A review. Sci. Robot. **3** (2018). https://doi.org/10.1126/scirobotics.aat5954
21. David, D., Thérouanne, P., Milhabet, I.: The acceptability of social robots: A scoping review of the recent literature. Comput. Hum. Behav. **137**, 107419 (2022). https://doi.org/10.1016/j.chb.2022.107419
22. Papakostas, G.A., Sidiropoulos, G.K., Papadopoulou, C.I., Vrochidou, E., Kaburlasos, V.G., Papadopoulou, M.T., Holeva, V., Nikopoulou, V.-A., Dalivigkas, N.: Social Robots in Special

Education: A systematic review. Electronics. **10**, 1398 (2021). https://doi.org/10.3390/electronics10121398
23. Zeng, F., Gan, W., Wang, Y., Liu, N., Yu, P.S.: Large Language Models for Robotics: A Survey. arXiv Prepr. arXiv:2311, (2023)
24. Fan, H., Liu, X., Fuh, J.Y.H., Lu, W.F., Li, B.: Embodied intelligence in manufacturing: Leveraging large language models for autonomous industrial robotics. J. Intell. Manuf. (2024). https://doi.org/10.1007/s10845-023-02294-y
25. National Center for Educational Statistics (NCES): Condition of Education: Annual Reports. Students with Disabilities. https://nces.ed.gov/programs/coe/indicator/cgg/students-with-disabilities
26. UNESCO Internation Insitute for Educational Planning: Special needs education. https://policytoolbox.iiep.unesco.org/glossary/special-education/
27. United Nations: Article 24 - Education. https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/article-24-education.html
28. Georgieva-Tsaneva, G., Andreeva, A., Tsvetkova, P., Lekova, A., Simonska, M., Stancheva-Popkostadinova, V., Dimitrov, G., Rasheva-Yordanova, K., Kostadinova, I.: Exploring the potential of Social Robots for Speech and Language Therapy: A review and analysis of interactive scenarios. Machines. **11**, 693 (2023). https://doi.org/10.3390/machines11070693
29. Francis, G., Deniz, E., Torgerson, C., Toseeb, U.: Play-based interventions for mental health: A systematic review and meta-analysis focused on children and adolescents with autism spectrum disorder and developmental language disorder. Autism Dev. Lang. Impair. **7**, 239694152110731 (2022). https://doi.org/10.1177/23969415211073118
30. Neeharika, C.H., Riyazuddin, Y.M.: Artificial Intelligence in Children with Special Need Education. In: 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT). pp. 519–523. IEEE (2023)
31. Nguyen, T.: Leveraging large Language models for Educational Enhancement: A Case Study of ChatGPT, BingChat, and Bard. Preprints. **2023091554** (2023). https://doi.org/10.20944/preprints202309.1554.v1
32. Tlili, A., Lin, V., Chen, N.S., Huang, R.: Kinshuk: A systematic review on robot-assisted special education from the activity theory perspective. Educ. Technol. Soc. **23**, 95–109 (2020)
33. Elhajoui, A., Alaoui, O.Y., El Kharki, O., Wahbi, M., Boulassal, H., Maatouk, M.: ChatGPT for a Flexible Higher Education: A Rapid Review of the Literature. In: Lecture Notes in Networks and Systems. pp. 369–375 (2024)
34. Marino, M.T., Vasquez, E., Dieker, L., Basham, J., Blackorby, J.: The future of Artificial Intelligence in Special Education Technology. J. Spec. Educ. Technol. **38**, 404–416 (2023). https://doi.org/10.1177/01626434231165977
35. Naveed, H., Khan, A.U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., Mian, A.: A Comprehensive Overview of Large Language Models. arXiv Prepr. arXiv:2307, (2023)
36. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. **5**, 115–133 (1943). https://doi.org/10.1007/BF02478259
37. Hutchins, J.: The First Public Demonstration of Machine Translation: The Georgetown-IBM System, 7th January 1954. MT News Int (1994)
38. Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM. **9**, 36–45 (1966). https://doi.org/10.1145/365153.365168
39. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**, 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
40. Mikolov, T., Kombrink, S., Burget, L., Cernocky, J., Khudanpur, S.: Extensions of recurrent neural network language model. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 5528–5531. IEEE (2011)
41. Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., McClosky, D.: The Stanford CoreNLP Natural Language Processing Toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. pp. 55–60. Association for Computational Linguistics, Stroudsburg, PA, USA (2014)
42. Helms, M., Ault, S.V., Mao, G., Wang, J.: An Overview of Google Brain and Its Applications. In: Proceedings of the 2018 International Conference on Big Data and Education. pp. 72–75. ACM, New York, NY, USA (2018)
43. Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., Dean, J.: Google's neural machine translation system: Bridging the gap between Human and Machine Translation. arXiv Prepr. arXiv:1609.08144, (2016)
44. Ahmed, K., Keskar, N.S., Socher, R.: Weighted Transformer Network for Machine Translation. arXiv Prepr. arXiv:1711.02132, (2017)
45. Radford, A., Salimans, T.: Improving Language Understanding by Generative Pre-Training (transformer in real world). OpenAI. (2018)
46. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional transformers for Language understanding. arXiv Prepr. arXiv:1810.04805, (2018)
47. Radford Alec, W., Rewon, J.C., David, L.: Amodei Dario, Sutskever Ilya: Language models are unsupervised Multitask learners| enhanced reader. OpenAI Blog. **1**, 9 (2019)
48. Shoeybi, M., Patwary, M., Puri, R., LeGresley, P., Casper, J., Catanzaro, B.: Megatron-LM: Training Multi-billion Parameter Language models using model parallelism. arXiv Prepr. arXiv:1909.08053, (2019)
49. Adamopoulou, E., Moussiades, L.: Chatbots: History, technology, and applications. Mach. Learn. Appl. **2**, 100006 (2020). https://doi.org/10.1016/j.mlwa.2020.100006
50. Adamopoulou, E., Moussiades, L.: An Overview of Chatbot Technology. In: IFIP international conference on Artificial Intelligence Applications and Innovations. pp. 373–383 (2020)
51. Thapa, S., Adhikari, S.: ChatGPT, Bard, and Large Language Models for Biomedical Research: Opportunities and pitfalls. Ann. Biomed. Eng. **51**, 2647–2651 (2023). https://doi.org/10.1007/s10439-023-03284-0
52. Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C.C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., Fuller, B., Gao, C., Goswami, V., Goyal, N., Hartshorn, A., Hosseini, S., Hou, R., Inan, H., Kardas, M., Kerkez, V., Khabsa, M., Kloumann, I., Korenev, A., Koura, P.S., Lachaux, M.-A., Lavril, T., Lee, J., Liskovich, D., Lu, Y., Mao, Y., Martinet, X., Mihaylov, T., Mishra, P., Molybog, I., Nie, Y., Poulton, A., Reizenstein, J., Rungta, R., Saladi, K., Schelten, A., Silva, R., Smith, E.M., Subramanian, R., Tan, X.E., Tang, B., Taylor, R., Williams, A., Kuan, J.X., Xu, P., Yan, Z., Zarov, I., Zhang, Y., Fan, A., Kambadur, M., Narang, S., Rodriguez, A., Stojnic, R., Edunov: S., Scialom, T.: (2023). Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv Prepr. arXiv:2307
53. Almazrouei, E., Alobeidli, H., Alshamsi, A., Cappelli, A., Cojocaru, R., Debbah, M., Goffinet, É., Hesslow, D., Launay, J., Malartic, Q., Mazzotta, D., Noune, B., Pannier, B., Penedo, G.: The Falcon Series of Open Language Models. arXiv Prepr. arXiv:2311, (2023)

54. Schiavo, F., Campitiello, L., Todino, M.D., Di Tore, P.A.: Educational Robots, Emotion Recognition and ASD: New Horizon in Special Education. Educ. Sci. **14**, 258 (2024). https://doi.org/10.3390/educsci14030258
55. Alemi, M., Meghdari, A., Ghazisaedy, M.: The Impact of Social Robotics on L2 Learners' anxiety and attitude in English Vocabulary Acquisition. Int. J. Soc. Robot. **7**, 523–535 (2015). https://doi.org/10.1007/s12369-015-0286-y
56. Kaburlasos, V.G., Vrochidou, E.: Social Robots for Pedagogical Rehabilitation: Trends and Novel Modeling Principles. In: Global., I. (ed.) Cyber-Physical Systems for Social Applications. pp. 1–12 (2019)
57. Gordon, G., Spaulding, S., Kory Westlund, J., Lee, J.J., Plummer, L., Martinez, M., Das, M., Breazeal, C.: Affective Personalization of a Social Robot Tutor for Children's Second Language Skills. Proc. AAAI Conf. Artif. Intell. 30, (2016). https://doi.org/10.1609/aaai.v30i1.9914
58. VanLEHN, K.: The relative effectiveness of human tutoring, Intelligent Tutoring Systems, and other Tutoring systems. Educ. Psychol. **46**, 197–221 (2011). https://doi.org/10.1080/00461520.2011.611369
59. Cano, S., Díaz-Arancibia, J., Arango-López, J., Libreros, J.E., García, M.: Design path for a Social Robot for Emotional Communication for Children with Autism Spectrum disorder (ASD). Sensors. **23**, 5291 (2023). https://doi.org/10.3390/s23115291
60. Sonderegger, S., HOW GENERATIVE LANGUAGE, MODELS CAN ENHANCE INTERACTIVE LEARNING WITH SOCIAL: ROBOTS. In: Proceeedings of the 19th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2022). IADIS Press (2022)
61. Bhat, S., Nguyen, H.A., Moore, S., Stamper, J., Sakr, M., Nyberg, E.: Towards automated generation and evaluation of questions in educational domains. In: Proceedings of the 15th International Conference on Educational Data Mining (Vol. 701) (2022)
62. Kitchenham, B.: Procedures for performing systematic reviews. Keele, UK, Keele University (2004)
63. Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D.: The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. **372**, n71 (2021). https://doi.org/10.1136/bmj.n71
64. Schotten, M., el Aisati, M., Meester, W.J.N., Steiginga, S., Ross, C.A.: A Brief History of Scopus: The World's Largest Abstract and Citation Database of Scientific Literature. In: Research Analytics. pp. 31–58. Auerbach Publications, Boca Raton, FL: Taylor & Francis, (2017) (2018)
65. Rane, N.: Enhancing the quality of teaching and learning through ChatGPT and similar large Language models: Challenges, future prospects, and ethical considerations in Education. SSRN Electron. J. (2023). https://doi.org/10.2139/ssrn.4599104
66. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., Kasneci, G.: ChatGPT for good? On opportunities and challenges of large language models for education. Learn. Individ Differ. **103**, 102274 (2023). https://doi.org/10.1016/j.lindif.2023.102274
67. Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y.-S., Gašević, D., Chen, G.: Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT. In: 2023 IEEE International Conference on Advanced Learning Technologies (ICALT). pp. 323–325. IEEE (2023)
68. Young, J.C., Shishido, M.: Investigating OpenAI's ChatGPT potentials in Generating Chatbot's dialogue for English as a Foreign Language Learning. Int. J. Adv. Comput. Sci. Appl. **14** (2023). https://doi.org/10.14569/IJACSA.2023.0140607
69. Bonner, E., Lege, R., Frazier, E.: LARGE, LANGUAGE MODEL-BASED ARTIFICIAL, INTELLIGENCE IN THE LANGUAGE CLASSROOM: PRACTICAL IDEAS FOR TEACHING. Teach:. English With Technol. (2023). (2023). https://doi.org/10.56297/BKAM1691/WIEO1749
70. Caines, A., Benedetto, L., Taslimipoor, S., Davis, C., Gao, Y., Andersen, O., Yuan, Z., Elliott, M., Moore, R., Bryant, C., Rei, M., Yannakoudakis, H., Mullooly, A., Nicholls, D., Buttery, P.: On the application of large Language models for language teaching and assessment technology. arXiv Prepr. arXiv:2307.08393, (2023)
71. Peng, Y., Nabae, H., Funabora, Y., Suzumori, K.: Peristaltic transporting device inspired by large intestine structure. Sens. Actuators Phys. **365**, 114840 (2024). https://doi.org/10.1016/j.sna.2023.114840
72. Peng, Y., Nabae, H., Funabora, Y., Suzumori, K.: Controlling a peristaltic robot inspired by inchworms. Biomim. Intell. Robot. **4**, 100146 (2024). https://doi.org/10.1016/j.birob.2024.100146
73. Peng, Y., Zhang, C., Hu, C., Mao, Z.: Integration of Large Language Models in Control of EHD Pumps for Precise Color Synthesis. arXiv Prepr. arXiv:2401, (2024)
74. Hadi, M.U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M.B., Mirjalili, S.: Large language models: A comprehensive survey of its applications, challenges, limitations, and future prospects. Authorea Prepr. (2023). https://doi.org/10.36227/techrxiv.23589741.v4
75. Jeon, J., Lee, S.: Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. Educ. Inf. Technol. **28**, 15873–15892 (2023). https://doi.org/10.1007/s10639-023-11834-1
76. Raiaan, M.A.K., Mukta, M.S.H., Fatema, K., Fahad, N.M., Sakib, S., Mim, M.M.J., Azam, S.: A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. TechRxiv. (2023). https://doi.org/10.36227/techrxiv.24171183.v1
77. Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., Gašević, D.: Practical and ethical challenges of large Language models in Education: A systematic scoping review. arXiv Prepr arXiv. 2303 (2023). https://doi.org/10.1111/bjet.13370
78. Zhang, C., Chen, J., Li, J., Peng, Y., Mao, Z.: Large language models for human–robot interaction: A review. Biomim. Intell. Robot. **3**, 100131 (2023). https://doi.org/10.1016/j.birob.2023.100131
79. Bottega, J.A., Kich, V.A., de Jesus, J.C., Steinmetz, R., Kolling, A.H., Grando, R.B., Guerra, R. da, Gamarra, S.: D.F.T.: Jubileo: An Immersive Simulation Framework for Social Robot Design. J. Intell. Robot. Syst. 109, 91 (2023). https://doi.org/10.1007/s10846-023-01991-3
80. Ye, Y., You, H. and Du, J.: Improved Trust in Human-Robot Collaboration With ChatGPT," IEEE Access 11, 55748–55754 (2023)
81. Murali, P., Steenstra, I., Yun, H.S., Shamekhi, A., Bickmore, T.: Improving Multiparty Interactions with a Robot Using Large Language Models. In: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–8. ACM, New York, NY, USA (2023)
82. Billing, E., Rosén, J., Lamb, M.: Language Models for Human-Robot Interaction. In: Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. pp. 905–906. ACM, New York, NY, USA (2023)

83. Axelsson, A., Skantze, G.: Do You Follow? A Fully Automated System for Adaptive Robot Presenters. In: Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. pp. 102–111. ACM, New York, NY, USA (2023)
84. Mishra, C., Verdonschot, R., Hagoort, P., Skantze, G.: Real-time emotion generation in human-robot dialogue using large language models. Front. Robot AI. **10** (2023). https://doi.org/10.3389/frobt.2023.1271610
85. Lozano, E.A., Sánchez-Torres, C.E., López-Nava, I.H., Favela, J.: An Open Framework for Nonverbal Communication in Human-Robot Interaction. Presented at the (2023)
86. Onorati, T., Castro-González, Á., del Valle, J.C., Díaz, P., Castillo, J.C.: Creating Personalized Verbal Human-Robot Interactions Using LLM with the Robot Mini. Presented at the (2023)
87. Sevilla-Salcedo, J., Fernádez-Rodicio, E., Martín-Galván, L., Castro-González, Á., Castillo, J.C., Salichs, M.A.: Using large Language models to shape Social Robots' Speech. Int. J. Interact. Multimed Artif. Intell. **8**, 6 (2023). https://doi.org/10.9781/ijimai.2023.07.008
88. Irfan, B., Kuoppamäki, S.M., Skantze, G.: Between Reality and Delusion: Challenges of Applying Large Language Models to Companion Robots for Open-Domain Dialogues with Older Adults. preprint. (2023). https://doi.org/10.21203/rs.3.rs-2884789/v1
89. Khoo, W., Hsu, L.-J., Amon, K.J., Chakilam, P.V., Chen, W.-C., Kaufman, Z., Lungu, A., Sato, H., Seliger, E., Swaminathan, M., Tsui, K.M., Crandall, D.J., Sabanović, S.: Spill the Tea: When Robot Conversation Agents Support Well-being for Older Adults. In: Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. pp. 178–182. ACM, New York, NY, USA (2023)
90. Wang, Z., Reisert, P., Nichols, E., Gomez, R.: Ain't Misbehavin' - Using LLMs to Generate Expressive Robot Behavior in Conversations with the Tabletop Robot Haru. In: Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. pp. 1105–1109. ACM, New York, NY, USA (2024)
91. Jokinen, K., Wilcock, G.: Exploring a Japanese Cooking Database. In: Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. pp. 578–582. ACM, New York, NY, USA (2024)
92. Rosén, J., Lindblom, J., Lamb, M., Billing, E.: Previous experience matters: An in-person investigation of expectations in Human–Robot Interaction. Int. J. Soc. Robot. **16**, 447–460 (2024). https://doi.org/10.1007/s12369-024-01107-3
93. Borg, A., Parodis, I., Skantze, G.: Creating Virtual Patients using Robots and Large Language Models: A Preliminary Study with Medical Students. In: Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. pp. 273–277. ACM, New York, NY, USA (2024)
94. Addlesee, A., Cherakara, N., Nelson, N., Hernández García, D., Gunson, N., Sieińska, W., Romeo, M., Dondrup, C., Lemon, O.: A Multi-party Conversational Social Robot Using LLMs. In: Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. pp. 1273–1275. ACM, New York, NY, USA (2024)
95. Kim, C.Y., Lee, C.P., Mutlu, B.: Understanding Large-Language Model (LLM)-powered Human-Robot Interaction. In: Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. pp. 371–380. ACM, New York, NY, USA (2024)
96. Cho, Y., Kim, M., Kim, S., Kwon, O., Kwon, R.D., Lee, Y., Lim, D.: Evaluating the efficacy of interactive Language Therapy based on LLM for high-functioning autistic adolescent psychological counseling. arXiv Prepr. arXiv:2311.09243, (2023)
97. Islam, A., Biswas, M.R., Zaghouani, W., Belhaouari, S.B., Shah, Z.: Pushing Boundaries: Exploring Zero Shot Object Classification with Large Multimodal Models. In: 2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS). pp. 1–5. IEEE (2023)
98. Packer, B., Keates, S.: Designing AI Writing Workflow UX for Reduced Cognitive Loads. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 306–325 (2023)
99. Goodman, S.M., Buehler, E., Clary, P., Coenen, A., Donsbach, A., Horne, T.N., Lahav, M., MacDonald, R., Michaels, R.B., Narayanan, A., Pushkarna, M., Riley, J., Santana, A., Shi, L., Sweeney, R., Weaver, P., Yuan, A., Morris, M.R.: LaMPost: Design and Evaluation of an AI-assisted Email Writing Prototype for Adults with Dyslexia. In: Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility. pp. 1–18. ACM, New York, NY, USA (2022)
100. Rajagopal, A., Nirmala, V., Jebadurai, I.J., Vedamanickam, A.M., Kumar, P.U.: Design of Generative Multimodal AI Agents to Enable Persons with Learning Disability. In: International Cconference on Multimodal Interaction. pp. 259–271. ACM, New York, NY, USA (2023)
101. Fang, Y., Xu, Y., Teng, Z., Jiang, Z., Xiang, W.: SocializeChat: a GPT-based AAC Tool for Social Communication Through Eye Gazing. In: Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing. pp. 128–132. ACM, New York, NY, USA (2023)
102. Furhat, Robotics: Furhat robot, https://furhatrobotics.com/
103. : EVA, Platform, E.V.A.: https://eva-social-robot.github.io/
104. Salichs, M.A., Castro-González, Á., Salichs, E.: Mini: A New Social Robot for the Elderly. Int. J. Soc. Robot. **12**, 1231–1249 (2020). https://doi.org/10.1007/s12369-020-00687-0
105. : LuxAI, QTrobot: https://luxai.com/humanoid-social-robot-for-research-and-teaching/
106. SoftBank, R.: Pepper, https://us.softbankrobotics.com/pepper
107. ALDEBARAN: NAO. https://www.aldebaran.com/en/nao
108. Waisberg, E., Ong, J., Masalkhi, M., Zaman, N., Sarker, P., Lee, A.G., Tavakkoli, A.: Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impairments. Eye. (2023). https://doi.org/10.1038/s41433-023-02842-z
109. Rane, N.L., Tawde, A., Choudhary, S.P., Rane, J.: Contribution and performance of ChatGPT and other large Language models (LLM) for scientific and research advancements: A double-edged sword. Int. Res. J. Mod. Eng. Technol. Sci. **5**, 875–899 (2023). https://doi.org/https://www.https://doi.org/10.56726/IRJMETS45312
110. Lim, J., Sa, I., MacDonald, B., Ahn, H.S.: A Sign Language Recognition System with Pepper, Lightweight-Transformer, and LLM. arXiv Prepr. arXiv:2309, (2023)
111. Mishra, R., Welch, K.C.: Towards Scalable Robotic Intervention of Children with Autism Spectrum Disorder Using LLMs. arXiv Prepr (2024)
112. Moxie, M., Robot: https://moxierobot.com/products/ai-robot
113. Khalid, S.J., Ali, I.A.: Mind Controlled Educational Robotic Toys for Physically Disabled Children: A Survey. In: 2022 International Conference on Computer Science and Software Engineering (CSASE). pp. 348–354. IEEE (2022)
114. Keyirobot: Loona Smart Robot. https://keyirobot.com/products/loona
115. Anki Cozmo Robot: Cozmo Robot 2.0. https://ankicozmorobot.com/
116. Living.ai: EMO: The Coolest AI Desktop Pet with Personality and Ideas. https://living.ai/emo/
117. Engineered Arts: AMECA: The Future Face Of Robotics. https://www.engineeredarts.co.uk/robot/ameca/

118. Toolify.ai: Revolutionary GPT-5 Robot Unveiled by OpenAI Sends Shockwaves Across AI Industry! https://www.toolify.ai/ai-news/revolutionary-gpt5-robot-unveiled-by-openai-sends-shockwaves-across-ai-industry-562740
119. Imran, M.M., Chatterjee, P., Damevski, K.: Uncovering the causes of emotions in Software developer communication using zero-shot LLMs. arXiv Prepr (2023). arXiv:2312
120. Papakostas, G.A., Strolis, A.K., Panagiotopoulos, F., Aitsidis, C.N.: Social Robot Selection: A Case Study in Education. In: 2018 26th International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2018. pp. 1–4 (2018)
121. IEEE: IEEE Ethically Aligned Design: (2019)
122. European Comission: Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
123. Intersoft consulting: The EU General Data Protection Regulation (GDPR). https://gdpr-info.eu/
124. Consumer Privacy Act: Federal Consumer Online Privacy Rights Act (COPRA). https://www.consumerprivacyact.com/federal/
125. Huang, J.T., Wang, W., Li, E.J., Lam, M.H., Ren, S., Yuan, Y., Lyu, M.: On the humanity of conversational ai: Evaluating the psychological portrayal of llms. In: Twelfth International Conference on Learning Representations. (2023)
126. Ke, L., Tong, S., Cheng, P., Peng, K.: Exploring the frontiers of LLMs in psychological applications: A Comprehensive Review. arXiv Prepr. arXiv:2401.01519, (2024)
127. Pinto, P.H.R., Araujo, V.M.U.D., Ferreira Junior, C.D.S., Goulart, L.L., Aguiar, G.S., Beltrão, J.V.C., Lira, P.D.D., Mendes, S.J.F., Monteiro, F.D.L.V., Avelino, E.L.: Assessing the Psychological Impact of Generative AI on computer and Data Science Education: An exploratory study. Preprints. **2023120379** (2023). https://doi.org/10.20944/preprints202312.0379.v2
128. DiPaola, D., Charisi, V., Breazeal, C., Sabanovic, S.: Children's Fundamental Rights in Human-Robot Interaction Research. In: Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. pp. 561–566. ACM, New York, NY, USA (2023)
129. Sun, Y., Zhang, J., Xiong, Y., Zhu, G.: Data security and privacy in Cloud Computing. Int. J. Distrib. Sens. Networks. **10**, 190903 (2014). https://doi.org/10.1155/2014/190903