Review

Advancements in AI-driven multilingual comprehension for social robot interactions: An extensive review

Yanling Dong and Xiaolan Zhou*

School of Foreign Languages & Literature, Shandong University, Jinan 250000, China

* **Correspondence:** Email: xlzhou@sdu.edu.cn.

Abstract: In the digital era, human-robot interaction is rapidly expanding, emphasizing the need for social robots to fluently understand and communicate in multiple languages. It is not merely about decoding words but about establishing connections and building trust. However, many current social robots are limited to popular languages, serving in fields like language teaching, healthcare and companionship. This review examines the AI-driven language abilities in social robots, providing a detailed overview of their applications and the challenges faced, from nuanced linguistic understanding to data quality and cultural adaptability. Last, we discuss the future of integrating advanced language models in robots to move beyond basic interactions and towards deeper emotional connections. Through this endeavor, we hope to provide a beacon for researchers, steering them towards a path where linguistic adeptness in robots is seamlessly melded with their capacity for genuine emotional engagement.

Keywords: social robotic interaction; AI-enhanced language comprehension; multilingual robotics; semantic interlingual challenges; empathetic robotic communication

1. Introduction

Social robots, known for their human-friendly interactions, are becoming common in a myriad of domains, from healthcare and education to the comfort of our homes [1–4]. These robots are often designed with the dual purpose of executing functional tasks while also establishing a dynamic rapport through communication and interaction. At the heart of social robotics lies the principle of human-robot interaction (HRI), which has advanced significantly over the years, adapting to the complexities of human communication and behavior [5]. However, a critical component of this progress is the development and integration of language translation and understanding capabilities, which is foundational to a truly versatile and effective social robot [6, 7]. This advancement not only enables these robots to cross the barrier of language, facilitating interaction with humans in a more context-specific and nuanced manner, but also resonates with the multicultural and multilingual reality of our global society.

The need for effective and inclusive communication with robots, irrespective of language barriers, has never been more relevant [8].

However, as we strive to advance the frontier of social robot interaction, we are met with significant challenges. Primary among these is the inherent limitation of monolingual interaction. Many current social robots can only communicate in a single language, often English, which restricts their functionality and universal appeal. Expanding their linguistic repertoire is therefore a critical step towards ensuring more inclusive and effective interactions. While efforts have been made to incorporate multilingual capabilities into social robots, these attempts have been plagued with issues of translation inaccuracies [9]. Understanding the subtleties and nuances inherent in human language, such as idioms, metaphors or cultural references, can pose significant difficulties for AI systems and thus hamper effective HRI. Moreover, another hurdle in this journey is the difficulty of understanding context and human intent. Human communication is rarely devoid of context; the same set of words can convey drastically different meanings depending on the surrounding conversation and nonverbal cues. While humans are naturally adept at grasping such subtleties, replicating this skill in social robots proves challenging [10]. Additionally, understanding human intent extends beyond the spoken words. It requires the ability to decipher indirect communication, sarcasm and cultural nuances, areas that are yet to be fully explored in social robotics.

With these challenges in mind, the urgency and significance of further exploration into AI-based language translation and understanding of social robot interaction come into clear focus. In the face of our multicultural, multilingual world, the ability of robots to understand and interact in multiple languages can break down barriers and foster a more universal adoption of social robots across diverse cultural contexts [11]. For instance, consider the case of healthcare robots deployed in eldercare facilities where residents may come from diverse linguistic backgrounds. These robots could provide comfort, monitor health and even assist in therapeutic activities [12]. However, the effectiveness of such robots would be drastically limited if they could not understand or respond accurately to the multilingual needs of the residents. Moreover, in an educational setting, a social robot that understands the language and culture of the learners can offer personalized assistance, helping to bridge the educational gap in linguistically diverse classrooms. Our study, through this comprehensive review, seeks to address these challenges by systematically examining the existing literature in the field, identifying gaps in current knowledge and presenting opportunities for further research. This work is driven by the belief that advancements in AI-based language translation and understanding can fundamentally transform the interaction between humans and robots, paving the way for a future where social robots are not merely tools but companions capable of understanding and communicating with us in the language we speak, the way we speak it.

The remainder of this paper is structured as follows. Section 2 explores language translation and understanding methods. Section 3 provides an overview of social robot interaction, covering text-based and language translation-based human-robot interaction. Section 4 showcases cutting-edge applications of language translation and understanding in various domains, such as domestic assistance, education, customer service and cross-cultural collaboration. Then, Section 5 discusses current challenges and future directions in the field. Finally, in Section 6, we summarize our key findings, emphasize the impact of AI in machine translation and human-robot interaction and highlight opportunities for future research.

2. Language translation and understanding methods

2.1. Related natural language processing techniques

Figure 1 illustrates the process of speech interaction, encompassing speech input processing, language processing, dialogue management, and speech output processing. It begins with speech recognition (ASR) to convert speech into text, followed by machine translation (MT) and natural language understanding (NLU) for language processing. Dialogue management involves user intent recognition, contextual understanding, and response generation, enabling a deeper understanding of the user's input. Finally, text-to-speech synthesis (TTS) converts the response text into synthesized speech output, allowing the robot to effectively communicate its response to the user. Then, this section will provide detailed explanations of the following four parts: speech recognition, machine translation (MT), sentiment analysis, and natural language understanding and generation (NLU/NLG).
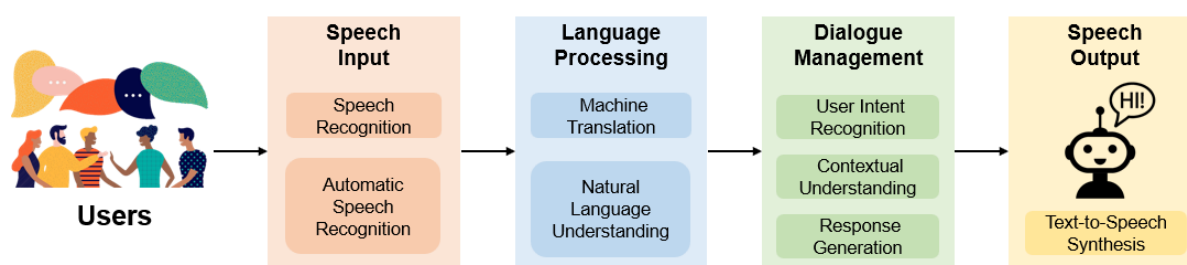

Figure 1. Natural language processing workflow.

2.1.1. Speech recognition

Speech recognition serves as a cornerstone for communication between humans and social robots. This technology's origin can be traced back to the 1960s, with IBM's Shoebox being one of the earliest systems capable of recognizing spoken digits and a limited set of words [13]. Over the decades, speech recognition technology has evolved significantly, with advancements fueled by machine learning and deep learning techniques. Microsoft's work in this arena, especially with products like Azure Speech Services, has contributed immensely to enhancing the accuracy and versatility of speech recognition systems across different applications, augmenting the linguistic capacities of social robots. In the context of social robot interaction, speech recognition forms the first line of processing in language translation and understanding [14]. It converts spoken language into written text, facilitating the robot's comprehension and subsequent response generation. This process encompasses challenges such as the diversity of human languages, accents and the presence of ambient noise, to name a few [15–17].

Advanced speech recognition algorithms, powered by deep learning methodologies, have shown remarkable capabilities in handling these challenges. For instance, Google's speech recognition system has been instrumental in providing accurate transcription services, even in noisy environments, paving the way for more efficient HRI. Similarly, Apple's Siri has been valuing robust speech recognition technology, enabling the assistant to understand and execute a wide range of user commands.

Furthermore, Microsoft's work in this arena, especially with products like Azure Speech Services, has contributed immensely to enhancing the accuracy and versatility of speech recognition systems across different applications, augmenting the linguistic capacities of social robots.

2.1.2. Machine translation

Machine translation (MT) has been a pivotal component in the evolution of HRI, enabling robots to understand and generate language beyond their programmed instructions [18]. Since its inception in the 1950s, machine translation has passed through several stages of development, from rule-based systems to statistical and, more recently, to neural network approaches [19–21]. Statistical machine translation (SMT) is a prominent example of the early stage of MT. Introduced in the late 1980s, SMT models rely on the analysis of bilingual text corpora to predict translations [22]. The introduction of neural machine translation (NMT) marked a significant leap forward. With its deep learning architecture, NMT provides end-to-end learning and can generate more natural and accurate translations. NMT models are capable of capturing the context and semantics of sentences, contributing to a more nuanced and effective translation [23]. In the context of social robot interaction, machine translation plays a crucial role in bridging the gap between different languages. Once the speech is recognized and converted into text, MT steps in to convert the text into a language that the robot understands. Subsequently, the robot's responses are translated back into the human user's language. The seamless integration of speech recognition and machine translation technologies enables social robots to communicate effectively and naturally with users of different languages [24].

2.1.3. Sentiment analysis

Sentiment analysis, also known as opinion mining, is a subfield of natural language processing (NLP) that identifies and extracts subjective information from source materials [25–27]. It is primarily used to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The technology has been broadly applied in text analysis, business analytics and social media monitoring. For social robots, sentiment analysis plays a crucial role in understanding human emotions, which is essential for effective HRI. By analyzing the sentiment of the user's input, social robots can adjust their responses accordingly, leading to more engaging and personalized interactions. For instance, if a user's input is detected as negative, the robot might respond in a way that shows empathy or attempts to uplift the user's mood [28].

2.1.4. Natural Language Understanding and Generation

Natural language understanding (NLU) and generation (NLG) are two critical aspects of NLP that deal with machine reading comprehension and the production of human-like text, respectively. NLU is the process of understanding and interpreting human language in a valuable way, which enables the social robot to understand and interpret the user's commands, questions and statements [29]. On the other hand, NLG is the task of converting information from computer databases or semantic intents into readable human language, which allows it to generate responses that are coherent, relevant and human-like [30, 31]. Together, NLU and NLG form the backbone of the conversational capabilities of social robots, enabling them to carry out meaningful and natural interactions with users.

2.2. *Case studies of AI-based translation and understanding*

The evolution of AI-based translation and understanding can be traced through several key applications that have significantly influenced the field. One of the earliest and most well-known is Google Translate. Launched in 2006, Google Translate initially used statistical machine translation, which translates text based on the analysis of bilingual text corpora. However, in 2016, Google introduced a neural machine translation system, which translates entire sentences at a time rather than piece by piece, providing more fluent and natural sounding translations [32]. Following Google Translate, iFlytek Translator made its debut. Developed by iFlytek, a Chinese information technology company, it is renowned for its high accuracy in speech recognition and translation, especially between English and Mandarin. The device uses deep learning technologies and can support translation between 50 languages [33].

In the realm of personal assistants, Apple's Siri is a notable example. Introduced in 2011, Siri uses natural language processing to interpret voice commands, answer questions, make recommendations and perform actions by delegating requests to a set of Internet services [34]. After that, Apple's Siri has been valuing robust language translation and understanding technology, enabling the assistant to understand and execute a wide range of user commands. Currently, Siri supports a multitude of languages and dialects, and can adapt to users' individual language usage and search preferences over time. Most recently, OpenAI's ChatGPT has emerged as a state-of-the-art language model. Trained on a diverse range of internet text, ChatGPT generates human-like text based on the input provided. It can translate languages, answer questions, write essays and even generate creative content like poetry [35, 36].

3. Overview of social robot interaction

3.1. *Text-based social robot interaction*

In the early days of computing, the primary mode of social robot interaction was through textual commands and responses. This form of interaction, although seemingly primitive by today's standards, laid the foundation for more complex forms of HRI that we see today [37–39]. One of the earliest examples of text-based social robot interaction is the ELIZA program developed by Weizenbaum at MIT in the 1960s [40]. ELIZA was a computer program that emulated a psychotherapist by using pattern matching and substitution methodology to simulate conversation. Despite its simplicity, ELIZA was able to demonstrate the illusion of understanding, which marked a significant milestone in the field of artificial intelligence and social robot interaction. However, text-based interaction lacks the richness of non-verbal cues, such as facial expressions and body language, which play a crucial role in human communication [41].

3.2. *Social robot interaction based on English speech*

The evolution of social robot interaction took a significant leap with the introduction of speech-based interaction. This development was largely facilitated by advancements in speech recognition technology, which allowed robots to understand and respond to spoken language. Compared to text-based social robot interaction, the interaction based on speech can provide a more natural and convenient interactive experience for users.

Among a variety of languages, English is the most often used language for social robot interac-

tion. Replika is a chatbot application designed to provide users with conversation and emotional support [42]. The initial version of it only supported English communication. In addition, many chatbots are used to enhance students' English language learning. In [43], Kanda et al. examined the potential for robots to form relationships with children and facilitate learning. A field trial at a Japanese elementary school involving English-speaking Robovie robots showed that initial interactions were frequent but declined over time. However, continued interaction during the second week predicted improvements in English skills, especially for children with prior proficiency or interest in English. Zakos [44] invented CLIVE, an artificially intelligent chatbot designed to facilitate English language learning through engaging and natural conversations. Unlike other tutoring systems, CLIVE offered an open and diverse range of topics, providing users with a lifelike and immersive language learning experience. In [45], Mini, a social robot, was designed to assist and accompany the elderly in various aspects of their daily lives. The robot offered services in personal assistance, entertainment, safety and stimulation, supporting cognitive and mental tasks.

However, while the use of English as the primary language for speech-based interaction has its advantages, such as a large user base and extensive research and resources, and also presents significant limitations. The primary limitation is the exclusion of non-English speakers, which constitutes a large portion of the global population [46]. This has led to a growing recognition of the need for multilingual capabilities in social robot interaction. Moreover, even within English speech-based interaction, there are challenges related to understanding accents, dialects and cultural nuances. This highlights the need for more advanced language understanding capabilities that can cater to the diversity of users.

3.3. Social robot interaction based on multiple languages

The development of language translation and understanding technologies has significantly broadened the scope of social robot interaction [47]. Unlike single-language-based virtual chatbots, most physical social robots are designed for multilingual interaction to cater to diverse user populations, thereby enhancing the user's understanding and engagement. This multilingual capability is particularly beneficial in multicultural and multilingual settings, where users may speak different languages.

One notable example of a social robot that leverages language translation and understanding is SoftBank's NAO robot, which has been used to teach English to native speakers of Dutch, German and Turkish as well as Dutch or German to Turkish-speaking children living in the Netherlands or Germany [48]. The NAO robot, through its ability to produce speech in various languages, provides a personalized, one-on-one tutoring experience that is both engaging and effective. Pepper, another social robot developed by SoftBank, utilizes natural language processing and speech recognition technologies, enabling it to recognize and understand text and speech inputs in multiple languages [49]. It supports various commonly used languages, including English, French, German, Italian, Spanish, Japanese and more. Therefore, users can communicate with Pepper in their familiar language of choice. In linguistically diverse L2 classrooms, social robots, which have been programmed to communicate in multiple languages, were used to assist L2 vocabulary learning [50]. Surprisingly, providing L1 translations through the robot did not demonstrate a facilitating effect on bilingual children's L2 word learning, contrary to initial predictions.

In recent years, the field of social robotics has witnessed significant advancements with the introduction of AI models like ChatGPT by OpenAI. ChatGPT, a large-scale language model, has been instrumental in enhancing the language translation and understanding capabilities of social robots. In

the military settings, ChatGPT is expected to play a role in various applications, including military robotics, battle space autonomy, automated target recognition and language translation [51]. Specifically, ChatGPT could be utilized for translating messages between several languages to improve understanding and communication between various military units or between the local communities and military in operational regions. In the industrial area, Ye et al. [52] investigated the impact of incorporating ChatGPT in a human-robot assembly task, where a robot arm controlled by RoboGPT assisted human operators. The study demonstrated that integrating ChatGPT significantly enhanced trust, attributed to improved communication and the robot's ability to understand and respond appropriately to human language. However, it is important to note that while ChatGPT represents a significant step forward, there are still challenges to overcome, such as ensuring the accuracy and appropriateness of its responses, and improving its ability to understand and respond to nonverbal cues.

In order to achieve natural and effective HRIs, social robots also need to possess the ability to perceive and identify complex emotional body language as well as display their own behaviors using similar communication modes [53, 54]. In [55], McColl and Nejat focused on the design of emotional body language for Brian 2.0, i.e., a human-like social robot, by incorporating various body postures and movements found in human emotion research. In a more recent study, Hong et al. [56] presented a novel multimodal emotional HRI architecture that combined body language and vocal intonation to detect user affect. Not only can the social robot interact with a human user in English but it can also determine its own emotional behavior based on user affect. For deaf and hard-of-hearing children, sign language plays a more critical role in their life rather than other oral languages. In [57], Meghdari et al. presented RASA (Robot Assistant for Social Aims), an educational social robot specifically designed for teaching Persian Sign Language (PSL) to children with hearing disabilities. RASA is characterized by its interactive social functionality, the ability to perform PSL with a dexterous upper-body and its cost-effectiveness. These examples demonstrate the fact that the integration of language translation and understanding technologies in social robots holds great promise for the future of HRI.

4. Cutting-edge applications of language translation and understanding in social robot interaction

In this section, we explore the forefront applications of AI-driven language translation and understanding in social robot interactions, spanning family assistance, educational support, service provision, travel guidance and cross-cultural collaboration. The documents referenced in this section have been meticulously chosen based on specific criteria. The majority of these were sourced from Google Scholar, while a minor fraction was curated from the broader internet, particularly from dedicated robot websites featuring news reports. It is imperative to highlight that the articles searched in Google Scholar underwent peer-review processes, ensuring their credibility and relevance. The content of these papers predominantly pertains to applications of robots equipped with multilingual comprehension and translation capabilities in HRIs, encompassing family, education, service, travel guide and cross-cultural collaboration domains. In terms of temporal relevance, the literature incorporated herein has been predominantly published within the last decade, with a significant emphasis on studies and advancements from the past five years. Figure 2 presents the various applications of social robot interaction based on language translation and understanding.
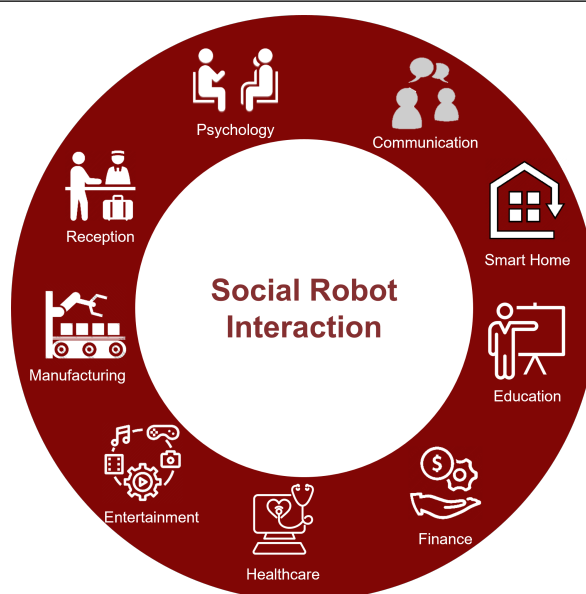
Figure 2. Various applications of social robot interaction based on language translation and understanding.

4.1. Assistive robots in family

In order to enhance overall family well-being and convenience, there is an emerging demand for intelligent family robots. These assistive robots, with outstanding language abilities and cute appearances, can assist with various aspects of family life, including companionship, communication, home automation and care support.

Assistive robots in the family are represented by companion chatbots. By extracting a formal meaning representation from natural language utterances, Atzeni and Atzori [58] proposed a language-independent approach for creating smart personal assistants and chatbots dubbed AskCo, which supported multiple languages. The system enables easy extensibility through Java source code and eliminates the need for training on large datasets, making it a flexible and efficient solution. Jelly, an AI-based chatbot developed using Facebook's Blenderbot, overcame language barriers by conversing with users in their native language, Nepali. It aimed to provide a comfortable and engaging conversation experience for those who struggle with English bots. The use of powerful text generation models enabled Jelly to understand romanized Nepali with English alphabets [59]. Regarding the Buddy robot [60], as shown in Figure 3, it is designed as an affordable family companion aimed at facilitating communication, ensuring home security, providing educational entertainment and even assisting with eldercare. Buddy is capable of autonomous movement and interacts with the environment through its integrated sensors, enabling object and facial recognition as well as language understanding and generation. It comes with pre-set languages of French and English and supports additional language downloads such as Japanese, Mandarin, Korean, etc.

Figure 3. Appearance design of Buddy [60].

With the increasing development of the Internet of Things (IoT) and home automation systems, there is a growing need for multilingual support to overcome language barriers, particularly for non-English speakers. In 2017, Eric et al. [61] explored the integration of voice control into a smart home automation system by leveraging voice recognition tools. Then the authors discussed different architectures for voice-enabled systems and evaluated available speech-to-text and text-to-speech engines, with a focus on the Google Cloud Speech API, which supported multi-language. Two years later, the smartphone is usually used to manage multiple remote controllers. Bajpai and Radha [62] proposed a solution using a smartphone and Arduino microcontroller to create a universal remote controller for cost-effective and convenient home automation. The study focused on developing a voice recognition system to control electronic appliances in a signal-based smart home network, enhancing the ease of use and accessibility for users. In 2021, Ta Multilanguage IoT Home Automation System was developed, specifically targeting elderly individuals in Malaysia, enabling them to control their home appliances using voice commands in their preferred language [63]. This research contributed to enhancing accessibility and usability for individuals with physical disabilities and older adults, who may face language challenges in utilizing smart home technologies. More recently, Soni et al. [64] introduced a novel approach to remotely control home appliances using smartphones, leveraging IoT technology. The system allows users to control appliances through voice commands in multiple languages, addressing the language barrier and enhancing system robustness and user convenience. The experimental results demonstrate a high level of performance, with an average success rate of 95.4%.

Smart and multipurpose voice recognition guiding robots have been developed to assist disabled people. For visually impaired individuals, Kalpana et al. [65] proposed an RTOS-enabled smart and multipurpose voice recognition guiding robot, which supported multiple regional languages. The robot, designed in the form of a dog, utilized Google Voice Recognition API to recognize user commands and employed light detection signals and a corner crossing algorithm for obstacle avoidance. Besides, It included a watchdog mode for abnormal movement detection and a self-charging feature using photovoltaic cells. A project aimed to develop a multi-language reading device to aid visually impaired

individuals in accessing information from regular books. The device utilized conversational AI technology, including image-to-text, translation and text-to-speech modules from Google Cloud. It supported multiple languages and could be used in public areas [66].

4.2. Assistive robots in education

Education applications are the primary source of the birth of multilingual social robots. Students from diverse linguistic backgrounds have different needs for language learning and educational experiences. The developed tutoring robots, especially NAO, can be applied in various educational aspects, including language tutoring, STEM training, metacognition tutoring, geometrical thinking training, oral proficiency development and facilitating communication and engagement in hybrid language classrooms. Therefore, developing various multilingual tutoring robots has become a trend. Table 1 illustrates the comparison of multilingual assistive robots in education applications.

Various studies have explored the effectiveness of social robots and mobile robots in language tutoring and STEM training. For example, Vogt et al. [67] implemented a large-scale experiment using a social robot to tutor young children in English vocabulary. Figure 4 illustrates the conic gestures the tutoring robot used. The robot, capable of translating between English and Dutch, was compared to a tablet application in terms of teaching new words. The results indicated that children were equally able to acquire and retain vocabulary from both the robot and the tablet. In another study, Leyzberg et al. [68] investigated the effectiveness of a personalization system for social robot tutors in a five-session English language learning task with native Spanish-speaking first-graders. The system, based on an adaptive Hidden Markov Model, ordered the curriculum to target individual skill proficiencies (Figure 5). The results demonstrated that participants who received personalized lessons from the robot tutor outperformed those who received non-personalized lessons. More recently, [69] explored the application of mobile robots in STEM training and proposed models that combined a mobile robot with an Android OS tablet for user interaction and voice control. They conducted experiments using an AI Processor to control the robot through voice commands in three languages (Korean, Vietnamese and English). The results showed high average confidence levels, providing a foundation for developing systems that support student learning through voice interaction with multi-language mobile robots. Furthermore, Schlippe et al. [70] developed a multilingual interactive conversational AI tutoring system for exam preparation. The system utilized a multilingual bidirectional encoder representations from transformers (M-BERT) model to automatically score free-text answers in 26 languages. It leveraged learning analytics, crowdsourcing and gamification to enhance the learning experience and adapt the system.

Table 1. Comparison of multilingual assistive robots in education applications. (processed by authors)

| Ref.	Robots	Languages	Subjects	Applications
[67]	NAO robot	English, Dutch	194 children	Tutoring children English vocabulary
[68]	Keepon robot	English, Spanish	First-graders	Personalized robot tutoring
[69]	Mobile robot	Korean, Vietnamese, English	N/A	STEM training; user interaction
[70]	Chatbot	26 languages	51 people	Exam preparation
[71]	NAO robot	English, German	40 participants	Tutoring foreign language words
[72]	NAO robot	Chinese, English	Preschoolers	Teaching preschoolers read and spell
[73]	NAO robot	English, Chinese, Japanese and Korean	19 college students	Individual tutoring and interactive learning experiences for students
[74]	NAO robot	Norwegian, English	20 children	Children's language learning progress in Norwegian day-care centers
[75]	NAO robot	Chinese, English	24 primary school students	English teaching
[76]	Telepresence robot	Finnish, German, Swedish and English	10–20 students	Classroom interaction; supporting the remote students
[77]	Telepresence robot	Japanese, English	more than 50 children	International communication between distant classrooms
[78]	EngSISLA	English, Hindi, Punjabi	different age group speakers	Translating the speech to Indian Sign Language
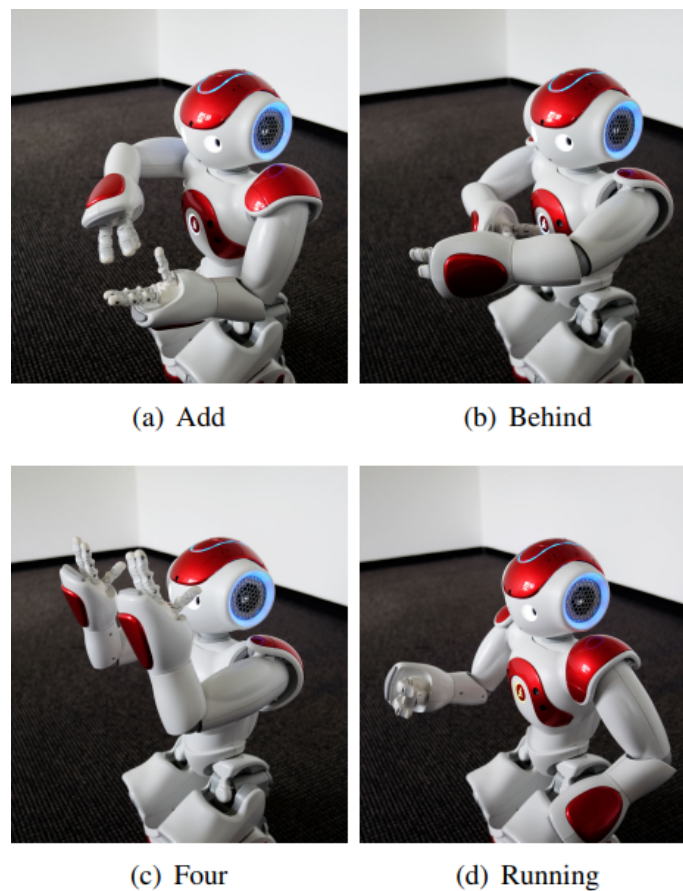
Figure 4. Examples of iconic gestures used in this study, photographed from the learner's perspective [67].

Figure 5. (a) A first-grade student interacts with the robot tutor. The caption here is an English translation of what the robot is saying in Spanish. (b) A Keepon robot [68].

Several studies have explored the use of NAO robots in language tutoring and educational settings, showcasing their potential to personalize tutoring, engage learners, enhance language proficiency and address educational challenges such as teacher shortages. In their novel approach, Schodde et al. [71] utilized a Bayesian knowledge tracing model combined with tutoring actions to personalize language tutoring in HRI, using word pairs in the artificial language Vimmi to prevent associations with known words or languages. Evaluation results demonstrated the superior effectiveness of the adaptive model in facilitating successful L2 word learning compared to randomized training. Another study by He et al. [72] focused on educational purposes and introduced a multi-language robot system that employed voice interaction and automatic questioning to engage learners in metacognition tutoring and geometrical thinking training. In the context of enhancing oral English proficiency, Lin et al. [73] developed the English oral training robot tutor system (EOTRTS), utilizing NAO, a social robot, to provide individual tutoring and interactive learning experiences for students in Taiwan. This system also had the potential to facilitate the learning of other foreign languages such as Japanese and Korean. Furthermore, a study in 2021 transformed the language shower program into a digital solution using a smartphone/tablet app and an NAO robot, demonstrating its positive impact on children's language learning progress in Norwegian day-care centers [74]. This highlights the potential of social robots in enhancing language learning. Addressing the shortage of English teachers in Taiwan, the modular English teaching multi-robot system (METMRS) employed NAO as the main teacher and Zenbo Junior robots as assistants, offering an innovative solution for English education [75].

Telepresence robots have emerged as valuable tools in educational settings, facilitating communication and engagement across language barriers and enhancing the learning experience for remote students. Jakonen and Jauni [76] explored the use of telepresence robots in hybrid language classrooms, where remote students participate through videoconferencing technology. Their findings highlight how telepresence robots enhance remote students' engagement and contribute to the multimodal meaning-making in hybrid language teaching. Similarly, Tanaka et al. [77] discussed the outcomes of a JST PRESTO project that utilized child-operated telepresence robots to facilitate international communication between classrooms, demonstrating the effectiveness of the system in enabling young children to communicate across language barriers.

Emerging technologies in education and communication, such as multilingual tutoring and speech-to-sign language translation systems, are transforming learning experiences and facilitating effective communication across language barriers. For example, Roybi, one of the most popular multilingual tutoring robots on the market, offers children an individualized educational experience through the use of AI. This interactive robot introduces children to technology, mathematics and science while engaging with them in various languages, including Spanish, French, English and Mandarin. In a similar field, a system called SISLA was proposed in [78], which utilizes a 3D avatar to translate speech into Indian Sign Language. With impressive accuracy rates of 91% for English and 89% for Punjabi and Hindi, usability testing confirms its effectiveness for educational and communication purposes, particularly for the hearing impaired.

4.3. Assistive robots in service

Driven by the need to improve customer experiences, various assistive robots in service have been developed. For instance, interactive information support systems and banking robots, have been designed to enhance concierge service. Additionally, assistive robots in healthcare, rehabilitation and

mental health have emerged as innovative solutions, providing support in hospitals, monitoring emotional well-being and improving accessibility for individuals with disabilities or elderly individuals. These cutting-edge applications demonstrate the transformative impact of language translation and understanding in social robot interaction across various domains.

To upgrade concierge services in hotels, Yamamoto et al. [79] proposed an interactive information support system utilizing smart devices and robot partners. The system comprises robot partners for communication and interaction with users and informationally structured space servers for data processing and personalized recommendations. It should be noted that the cute robots can select their communication language based on voice recognition through greeting. Its basic conversation flow is shown in Figure 6.
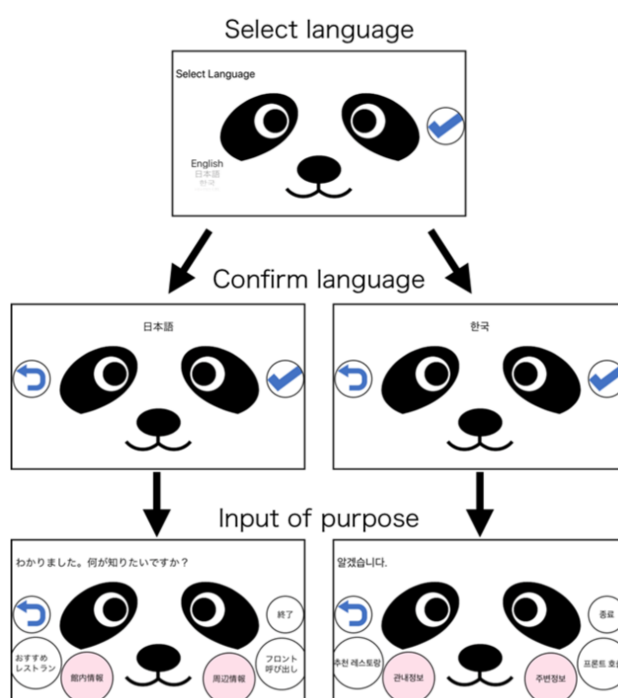

Figure 6. Scene transition [79].

Advancements in automatic speech recognition (ASR) and humanoid robot technologies are transforming the banking industry, enhancing customer experiences and overcoming language barriers. In a study conducted in Greece [80], researchers developed innovative methodologies for voice activity detection and noise elimination in budget robots, enabling effective ASR in challenging acoustically quasi-stationary environments. Furthermore, showcasing the potential of AI-driven robots in the banking sector, Pepper, a multi-linguistic humanoid robot, has made a positive impact at BBBank [81]. With its friendly and helpful demeanor, Pepper has assisted customers in various tasks, such as blocking stolen credit cards and providing relevant information. This successful integration of Pepper highlights its ability to enhance customer experience and illustrates the growing significance of robots in the banking industry.

Leveraging recent advancements in mobile speech translation and cognitive architectures, multilin-

gual promotional robots have emerged with great potential. In the pursuit of user-friendly and adaptable speech-to-speech translation systems for mobile devices, Yun et al. [82] developed a robust system by leveraging a large language and speech database. This research showcased the successful creation of a mobile-based, multi-language translation system capable of operating in real-world environments. Building upon this, Romero et al. [83] introduced the CORTEX cognitive architecture for social robots. By integrating different levels of abstraction into a unified deep space representation (DSR), this architecture facilitated agent interaction and behavior execution. The utilization of Microsoft's Kinect program in a separate embedded computer further enhanced the system's multi-language and multi-OS capabilities, exemplifying the potential of such technology in robotics.

In recent years, multilingual service robots have gained popularity, empowering diverse user groups with multimodal capabilities and extensive language support. Therefore, many innovative solutions are proposed. The PaeLife project, conducted in 2015, aimed to develop AALFred, a multimodal and multilingual virtual personal life assistant for senior citizens [84]. The project focused on various aspects, including collecting elderly speech corpora, optimizing speech recognition for elderly speakers, designing a reusable speech modality component and enabling automatic grammar translation to support multiple languages. After a few years, a software robot named Xiaoming was introduced. Xiaoming possessed multilingual and multimodal capabilities, allowing it to generate news, perform translation, read and animate avatars [85]. Voice cloning technology was utilized to synthesize speech in multiple languages, and Xiaomingbot achieved significant popularity on social media platforms by writing a substantial number of articles. Another notable research effort by Doumbouya et al. [86] addressed the challenge of providing speech recognition technology to illiterate populations. They explored unsupervised speech representation learning using noisy radio broadcasting archives and released datasets such as the West African Radio Corpus and West African Virtual Assistant Speech Recognition Corpus. Their work introduced the West African wav2vec speech encoder, which showed promising performance in multilingual speech recognition and language identification tasks.

In addition, according to [87], dependency on internet connectivity and language constraints hinder the effectiveness of smart assistants, such as Google Assistant, Siri and Alexa. To address these issues, a multilingual voice assistant system was developed using Raspberry Pi, enabling offline access to various languages and it allowed users to access information and perform tasks in their preferred language.

As the aging population grows and caregiver resources become limited, the demand for innovative technologies to assist and care for the elderly is on the rise. Socially assistive robots emerge as promising solutions for long-term elderly care. In 2016, Nuovo et al. [88] conducted an evaluation and development of a multi-modal user interface (MMUI) to enhance the usability and acceptance of assistive robot systems among elderly users. The experimental results demonstrated the effectiveness of the MMUI in improving flexibility and naturalness in interactions with the elderly. They also implemented multi-language speech recognition and text-to-speech (TTS) modules to facilitate communication using Nuance- and Acapela-VAAS respectively. Later, in 2018, a group of researchers further discussed the implementation of a user-friendly and acceptable service robotic system for the elderly, focusing on a web-based multi-modal user interface. Notably, it supported multi-language such as English, Italian and Swedish so as to enhance flexibility, naturalness, and acceptability of elderly-robot interaction [89]. In order to assist elderly individuals in adhering to their medication regimen, a novel robotic system was designed and evaluated using the NAO robot, which supported multi-language capabilities [90] (in Figure 7). This system utilized computer vision and a database to identify medication packaging,

detect the intended recipient, and ensure timely administration. Additionally, Giorgi et al. [91] enhanced HRI by developing human-like verbal and nonverbal behaviors in an NAO robot companion. It is worth noting that the robot served as a communicator in community activities with the elderly, offering multi-language translation capabilities through Cloud Services.
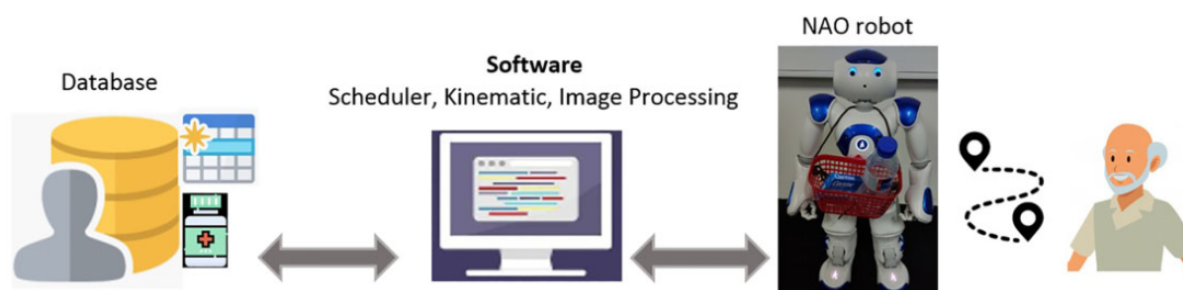

Figure 7. Robotic system overview (by: crisostomo) [90].

Advancements in voice-controlled robots and robust voice control systems have revolutionized the healthcare and rehabilitation sectors, providing potential support to hospitals and rehabilitation centers. In [92], Pramanik and his colleagues introduced a fully voice-controlled robot designed for hospitals, addressing staff overload and worker shortage situations. The robot's flexibility in movement, user-friendly characteristics and ability to accommodate diverse voices and languages make it suitable for satisfying the needs of both hospitals and patients. To meet the needs of patients with amputation, paralysis, and quadriplegia, a robust voice control system for rehabilitation robots was developed [93]. The system utilized advanced voice-recognition algorithms, such as hidden Markov model and dynamic time warping, to enhance accuracy and reduce errors (Figure 8). Its effectiveness in diverse noise environments and with multiple languages was demonstrated through validation experiments. In [94], the development of CLARA, a socially assistive robot (SAR), was discussed. Its appearance is presented in Figure 9. CLARA offers potential support for caregivers through its proactive, autonomous and adaptable nature. The integration of a multi-language interface using the Microsoft SDK enhances CLARA's perceptive and reactive capabilities, making it effective in various healthcare settings.
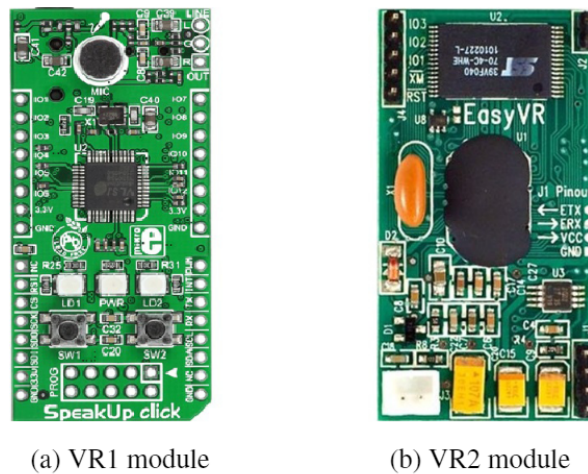
(a) VR1 module   (b) VR2 module

Figure 8. Voice-recognition modules implemented (by:Ruzaij) [93].
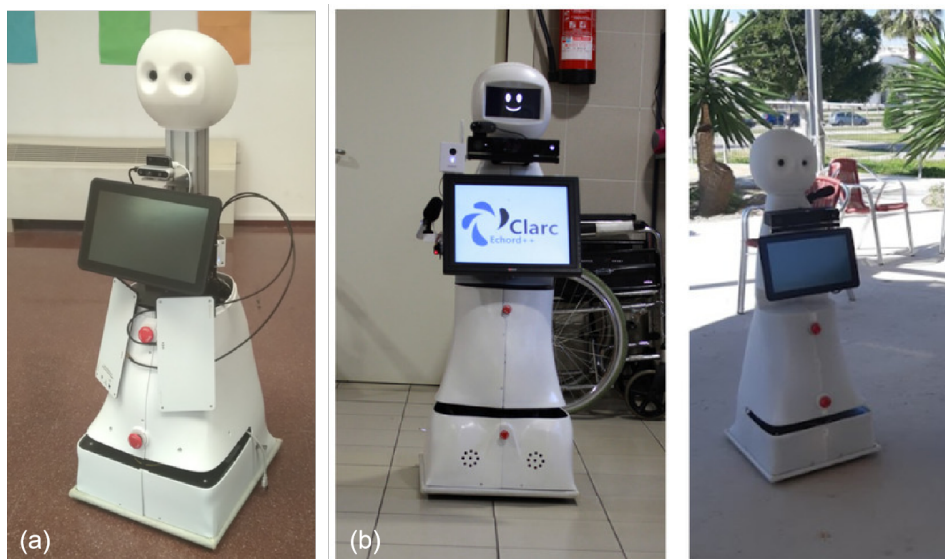

Figure 9. (a) One of the CLARA robots with the two RFID antennas. (b) External aspect of CLARA after adding (left) a first version and (right) the second version of the external housing [94].

Assistive technologies for individuals with physical disabilities, such as multi-input control systems and bilingual social networking service robots, improve accessibility and communication. These innovations empower users, enhancing their mobility and facilitating connections with peers and medical professionals. For instance, Ruzaij et al. [95] introduced a novel multi-input control system for rehabilitation applications, specifically designed for individuals with limited arm mobility. By integrating a voice controller unit and a head orientation control unit, the system employed various voice recognition algorithms and MEMS sensors to facilitate wheelchair control through user commands and head move-

ments. This hybrid voice controller not only enhanced voice recognition accuracy but also provided language flexibility, offering a promising solution for individuals with diverse needs. In a related context, Kobayashi et al. [96] proposed a bilingual social networking service (SNS) agency robot aimed at assisting individuals with physical disabilities in using tablets or smartphones for communication. Notably, this robot incorporated a voice user interface, enabling users to interact with others who share similar conditions or communicate with medical specialists in their native languages.

Recent advancements in healthcare robotics have introduced innovative solutions for evaluating mental health and monitoring emotional well-being in elderly individuals. In 2020, a multi-language robot interface that helped evaluate the mental health of elderly people through problem interaction was introduced by Yvanoff-Frenchin et al. [97], which was implemented on an embedded device for edge computing. The robot could interact with the user through appropriate language and it could process the answers and then, with the guidance of an expert, direct the questions and answers in the desired direction of treatment. At the same time, the device could filter out environmental noise and is suitable for placement anywhere in the home. In the same year, a robotic interface with multi-language capability was proposed [98] for monitoring and assessing the emotional health of elderly individuals through extended conversations. The system utilized voice interface and expert supervision to engage in automated conversations with clients, and the proposed method demonstrated compatibility with embedded platforms. One year later, Jibo, the social robot developed by NTT Disruption, has found application in the medical field as an empathetic healthcare companion [99] (Figure 10). Leveraging Microsoft Azure Cognitive Services, Jibo utilized AI capabilities to recognize people, understand moods and provide information and support for treatments. Notably, Jibo's multilingual communication abilities enable it to engage with patients in their preferred language, offering companionship, proactive assistance, video calling capabilities and reminders for treatment plans and exercises. During the COVID-19, Pepper has been deployed at Hořovice Hospital in the Czech Republic to assist in the fight against the pandemic [100]. With its ability to work tirelessly and be easily disinfected, Pepper helped enforce social distancing measures by detecting patients' temperatures and encouraging hand sanitization. The robot has been well-received by both staff and patients, improving the hospital experience and easing the burden on medical personnel.

Figure 10. Jibo robot for empathetic healthcare companion [99].

4.4. Assistive robots in travel guide

Nowadays, travel more and more takes the fancy of domestic and abroad tourists. Taking multiple factors into consideration, bright prospects lie in this field concerning assistive robots in travel guides.

In [101], a multi-lingual service system utilizing the iBeacon network for service robots was developed. By leveraging users' personal information stored in a dedicated app and the iBeacon region, the system enabled robots to understand users' language and provide personalized services. The collaborative nature of the system allowed for efficient resource utilization and had the potential to be applied in various domains, such as Olympic Game guidance. Additionally, Sun et al. [102] presented the "Concierge robot", a sightseeing support robot partner designed to recommend shops, restaurants and sightseeing spots to hotel visitors. The robot incorporated intelligent devices, a body and a four-wheel robot base, providing guide services through interactive multi-language utterances and a touch interface. Besides, Jeanpierre et al. [103] developed a robust system of autonomous robots that operated independently in complex public spaces, interacted with humans and assisted them in various environments. Equipped with a speech server with Microsoft Speech Recognition, the system demonstrates impressive effectiveness in interacting with visitors naturally. In 2019, Yoshiuchi et al. [104] explored the use of data analysis technology in service robot systems to improve business operations. By modifying service scenarios and analyzing collected data, the study demonstrated an 8.1% potential increase in business improvement, particularly in areas such as communication, image processing and multi-language processing.

Later, a voice-based attender robot with line-following capabilities and speech recognition was designed for university settings to assist with tasks such as passing circulars, interacting with parents and providing navigation assistance [105]. It can connect with humans through spoken natural language, specifically English and Kannada. The results verified the robot's effectiveness in facilitating communication with users, making it applicable not only to universities but also to other environments like railway stations, bus stations and factories. Recently, Zhang et al. [106] presented a voice control system for the LoCoBot WX250 robot, utilizing machine learning models and the BERT model for improved intent classification and keyword recognition. The system enhanced the interaction experience between humans and the robot, enabling it to act as a tour guide in museums. It could communicate with visitors via speaker and microphone, respond to instructions and even switch languages to accommodate foreign tourists. Pepper, another pioneering AI robot, revolutionized the tourism industry by breaking down language barriers and offering multi-linguistic communication and knowledge sharing [107] (Figure 11). With its touch of emotion and surprise, Pepper enhanced the museum experience, shared anecdotes and engaged visitors on a deeper level. Its interactive and proactive nature made it an invaluable tool for attracting and guiding visitors, creating a truly immersive and memorable cultural experience.

Figure 11. A guiding robot in library [107].

4.5. Assistive robots in cross-cultural collaboration

The development of assistive robots capable of speaking multiple languages in cross-cultural collaboration is essential for fostering effective communication and collaboration among individuals from diverse cultural backgrounds. These robots enable seamless interaction, understanding and cooperation between people who speak different languages, facilitating cross-cultural collaboration in various domains.

The integration of advanced technologies in robotics and AI is transforming industries, with applications ranging from industrial automation to agriculture. These innovations enhance productivity and efficiency, revolutionizing processes and addressing industry-specific challenges. Lin et al. [108] invented an automatic sorting system for industrial robots that integrates 3D visual perception, natural language interaction and automatic programming. Notably, the robot utilizes the open-source speech synthesis system (Ekho) for generating speech, supporting multiple languages and different platforms. In the same vein, Birch et al. [109] evaluated the effectiveness of a novel human-robot-interface for machine hole drilling, considering environmental factors on speech recognition accuracy. The developed speech recognition method, displayed in Figure 12, enabled HRI through a unique integration approach, employing DTW and distance comparison for word identification and language translation. Likewise, in [110], a mobile application that utilized AI and voice bot technology was developed to assist farmers in the agriculture sector. It featured a multi-linguistic voice bot for querying agricultural information and a suggestion bot for providing versatile suggestions related to weather, crops, fertilizers and soil. This AI-based system enhanced farming practices, increased agricultural production and addressed unknown issues faced by farmers.
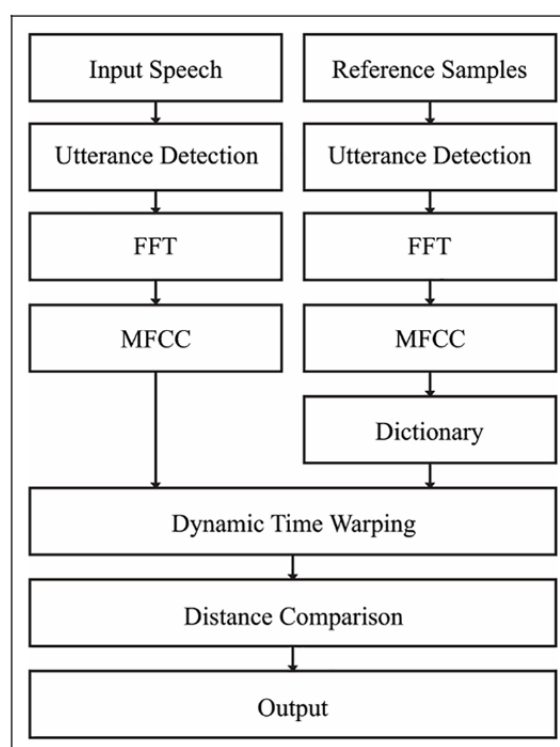
Figure 12. Algorithm flow chart (from: Birch) [109].

In a different context, Hong et al. [111] focused on implementing natural language-based communication between humans and fire fighting robots using ontological semantic technology (OST), which enabled a comprehensible understanding of meanings across multiple languages. The study expanded the application of OST to the domain of fire fighting, specifically addressing communication between robots and humans in Korean and English. To improve the office environment, a dialog agent that could understand natural language instructions from naive users was presented by Thomason et al. [112]. The agent incorporates a learning mechanism that induces training data from user paraphrases, enabling it to adapt to language variation without requiring large annotated corpora. Experimental results from web interfaces and a mobile robot deployed in an office environment demonstrated improved user satisfaction through learning from conversations. On top of that, Contreras et al. [113] explored the use of domain-based speech recognition to control drones in a more natural and human-like manner. By implementing an algorithm for command interpretation in both Spanish and English, the study demonstrated the effectiveness of voice instructions for drone control in a simulated domestic environment. The results showed improved accuracy in speech-to-action recognition, particularly with the phoneme matching approach, achieving high accuracy for both languages. In [114], a remote center of motion (RCM) based nasal robot was designed to assist in nasal surgery (Figure 13). Accordingly, a voice-based control method was proposed where surgeons provided direction instructions through the analysis of endoscopic images and a commercial speech recognition interface was used for offline grammar control, as shown in Figure 14. Additionally, a speech recognition interface was employed to create an offline grammar control word library that is compatible with both English and Chinese.
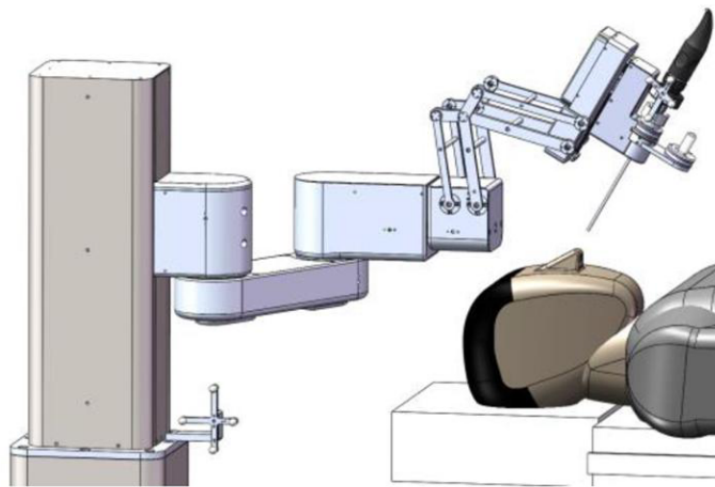
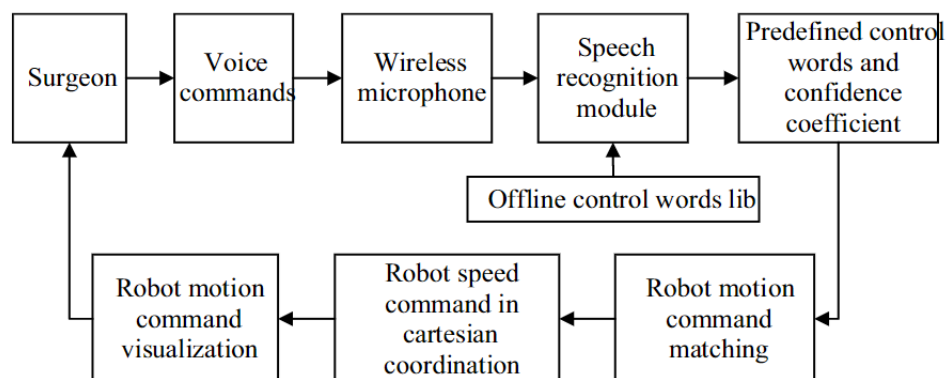**Figure 13.** The overall structure of the nasal endoscopic surgical robot [114].

Figure 14. Offline speech recognition process of robot motion instructions [114].

5. Recent challenges and future directions

5.1. Recent challenges

Although social robots are applied in many fields and have made technological breakthroughs in recent years, it still faces some challenges in language translation and understanding, as shown in the Figure 15.
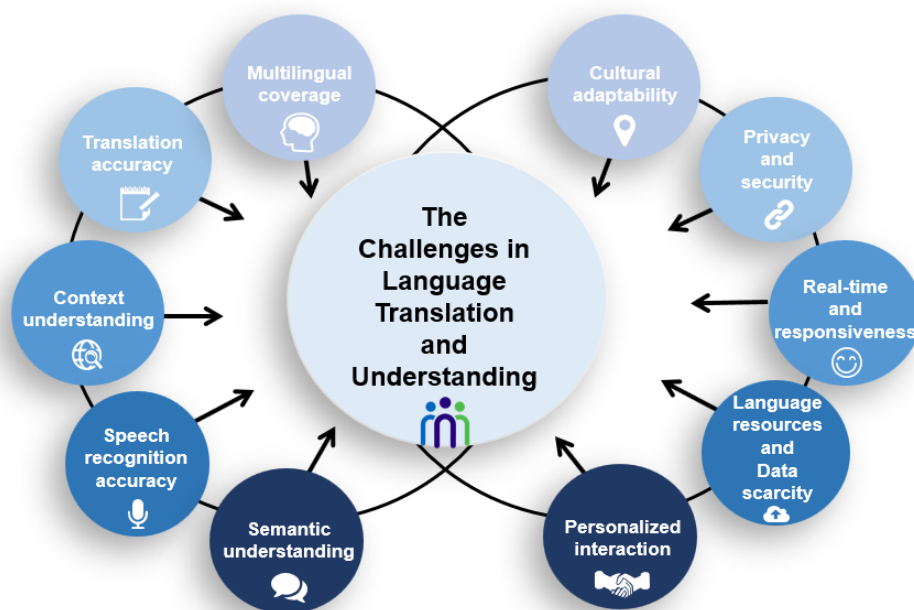
Figure 15. The challenges in language translation and understanding.

(1) Interlingual semantic understanding

Interlingual semantic understanding constitutes a critical aspect of AI-based language translation and understanding in social robot interaction. As robots are designed to communicate seamlessly with humans, their ability to understand semantics, not just literal translations, across multiple languages is crucial.

Interlingual semantic understanding typically involves techniques such as neural machine translation (NMT), where the system learns to translate by being trained on large amounts of text in both the source and target language. Moreover, models like BERT and GPT have enhanced semantic understanding by emphasizing the context of words. These models leverage deep learning and large-scale language representation to understand the semantics in one language and then generate the appropriate semantics in the target language.

There are two difficulties associated with interlingual semantic understanding. One major challenge is the issue of word sense disambiguation, differentiating the meaning of a word based on context. This becomes particularly challenging when a word or phrase in one language has multiple meanings in another. Additionally, understanding and correctly translating idioms, metaphors or cultural references is a formidable task for AI systems. These linguistic features often do not have direct equivalents across languages and require a deep understanding of both languages' cultures.

(2) Data scarcity and quality

Given that machine learning models are heavily reliant on large, high-quality datasets for training, the scarcity and inferior quality of data can pose substantial impediments to their performance.

Specific challenges can be concluded as follows. Current status reveals an uneven distribution of data across different languages. While extensive, high-quality datasets exist for popular languages like

English, many minority languages suffer from severe data scarcity. The consequence is an inherent bias in AI systems towards languages for which abundant data is available, resulting in less accurate translation and understanding capabilities for underrepresented languages. Furthermore, when ample data is available, its quality, including accuracy, consistency and relevance, might be compromised. For instance, training data may contain errors, be inconsistently annotated or simply not be representative of the diversity and complexity of real-world language use.

Furthermore, developing techniques to improve AI performance even with scarce or lower-quality data is a promising research direction. This includes methods such as transfer learning, where a pre-trained model on a large dataset can be fine-tuned on a smaller, specific dataset and data augmentation techniques to synthetically expand existing datasets.

Therefore, subsequent research is expected to develop techniques that can improve AI performance even with scarce or lower-quality data. This includes methods such as transfer learning, where a pre-trained model on a large dataset can be fine-tuned on a smaller, specific dataset and data augmentation techniques to synthetically expand existing datasets.

(3) Cultural adaptability and diversity

As social robots are envisioned to operate in multicultural societies and interact with people from different cultural backgrounds, their ability to adapt to various cultural norms and understand cultural diversity is essential.

The challenges associated with cultural adaptability and diversity are multifaceted. Language is deeply rooted in culture, carrying idiomatic expressions and metaphors that might be culturally exclusive. Another challenge is cultural bias. Training data used for AI systems often reflects the cultural characteristics of the regions where the data is sourced, which can inadvertently lead to cultural biases in AI models. Such biases could manifest as AI systems performing better for certain cultures while struggling with others. Furthermore, social etiquette, norms and expectations vary widely across different cultures. Designing social robots that can adapt to such a wide range of cultural expectations is an intricate challenge. Hence, addressing these challenges necessitates an interdisciplinary approach, combining insights from linguistics, anthropology, sociology and AI.

5.2. Future directions

5.2.1. Integration of language models

The GPT language model, particularly its latest iterations such as GPT-3 and GPT-4, has greatly impacted the field of language translation and understanding. Hence, the positive impact of GPT in the realm of social human interaction is evident. By providing social robots with the ability to understand and generate human-like responses, GPT has facilitated more nuanced and meaningful interactions. For instance, Nishihara et al. [115] developed an online algorithm for robots to acquire knowledge of natural language and object concepts by connecting recognized words to concepts. The model took into account the interdependence of words and concepts, enabling the robot to develop a more accurate language model and object concepts through unsupervised word segmentation and multimodal information. He and Mary [116], by reviewing the principles of ChatGPT, analyzed various aspects of robot perception and intelligence, excluding intrapersonal intelligence and proposed a multimodal approach using GPT-3 to implement seven types of robot intelligence. The proposed framework, called RobotGPT, paving the way for smarter robotic systems.

5.2.2. Overcoming the language barrier: multilingual and dialect understanding

Currently, many social robots are still largely restricted to English interaction. Those that do support multilingual interactions are often confined to specific domains such as language teaching, healthcare or social companionship, leaving broader applications, particularly work-related collaborations, underexplored. Looking ahead, a future where social robots shatter this language barrier, mastering not only multiple languages but also understanding dialects, is an exhilarating prospect. In [117], it suggests that a robot communicating in regional dialects or using a relaxed conversation style might be more warmly received. Andrist et al. [118] explored the impact of language and cultural context on the credibility of robot speech. Comparing Arabic-speaking robots in Lebanon and English-speaking robots in the USA, it revealed cultural differences in the importance of rhetorical cues and practical knowledge. These findings informed the design of culturally sensitive HRIs, particularly in relation to dialect usage.

5.2.3. From AI assistants to empathetic companions

Presently, interactions with social robots are primarily command-based, with the robots responding to explicit instructions from users. In the future, we envision social robots evolving from mere AI assistants to empathetic companions that truly understand human emotions, needs and desires, creating meaningful and enriching interactions. Imagine a scenario where you return home from a stressful day at work. Instead of merely offering to perform its usual tasks, it suggests relaxing activities like playing soothing music or initiating a calming meditation session. In a different scenario, let us imagine a tutoring robot assisting in a classroom. Beyond just answering questions or teaching language, the robot could gauge the understanding level of students by their facial expressions, confusion in their voice, or hesitation in their answers. It could then adjust the teaching speed or method to better accommodate the students' learning pace. In summary, the future of social robots lies in moving beyond command-based interaction to truly understanding and empathizing with human users.

6. Conclusions

In this literature review, we have explored the progression and current state of language translation and understanding in social robots, focusing particularly on the areas of multilingual capabilities and application in diverse domains. Our primary finding is that while social robots have shown promise in their ability to interact in one or two languages, there are still significant deficiencies, especially when it comes to broad multilingual interactions. Additionally, the application of multilingual social robots is mainly limited to areas like language teaching, healthcare and social companionship, with less prevalent use in sectors such as smart manufacturing or robot-assisted surgery.

This review provides a comprehensive look at the advancements made in the past decade from the perspective of social robot applications. We have detailed the current challenges faced in this domain, including interlingual semantic understanding, data scarcity and quality and cultural adaptability and diversity. By outlining these challenges, we hope to contribute to the research field by identifying the areas in need of focus and further development.

In conclusion, this literature review captures the evolution of language translation and understanding in social robots, summarizing the major challenges faced and outlining a roadmap for future research

directions. As we continue to advance in AI and robotics, we expect that this review will serve as a reference point for subsequent research aimed at enhancing the multilingual capabilities and empathetic interactions of social robots.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. O. Mubin, J. Henderson, C. Bartneck, You just do not understand me! Speech recognition in human robot interaction, in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, (2014), 637–642. https://doi.org/10.1109/ROMAN.2014.6926324

2. T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, F. Tanaka, Social robots for education: a review, *Sci. Rob.*, **3** (2018), eaat5954. https://doi.org/10.1126/scirobotics.aat5954

3. Y. Wang, S. Zhong, G. Wang, Preventing online disinformation propagation: cost-effective dynamic budget allocation of refutation, media censorship, and social bot detection, *Math. Biosci. Eng.*, **20** (2023), 13113–13132. https://doi.org/10.3934/mbe.2023584

4. C. A. Cifuentes, M. J. Pinto, N. Céspedes, M. Múnera, Social robots in therapy and care, *Curr. Rob. Rep.*, **1** (2020), 59–74. https://doi.org/10.1007/s43154-020-00009-2

5. H. Su, W. Qi, J. Chen, D. Zhang, Fuzzy approximation-based task-space control of robot manipulators with remote center of motion constraint, *IEEE Trans. Fuzzy Syst.*, **30** (2022), 1564–1573. https://doi.org/10.1109/TFUZZ.2022.3157075

6. J. Hirschberg, C. D. Manning, Advances in natural language processing, *Science*, **349** (2015), 261–266. https://doi.org/10.1126/science.aaa8685

7. S. H. Paplu, K. Berns, Towards linguistic and cognitive competence for socially interactive robots, in *Robot Intelligence Technology and Applications 6*, Springer, (2021), 520–530. https://doi.org/10.1007/978-3-030-97672-9_47

8. E. B. Onyeulo, V. Gandhi, What makes a social robot good at interacting with humans? *Information*, **11** (2020), 43. https://doi.org/10.3390/info11010043

9. C. Ke, V. W. Lou, K. C. Tan, M. Y. Wai, L. L. Chan, Changes in technology acceptance among older people with dementia: the role of social robot engagement, *Int. J. Med. Inf.*, **141** (2020), 104241. https://doi.org/10.1016/j.ijmedinf.2020.104241

10. Y. Kim, H. Chen, S. Alghowinem, C. Breazeal, H. W. Park, Joint engagement classification using video augmentation techniques for multi-person human-robot interaction, preprint, arXiv:2212.14128.

11. A. A. Allaban, M. Wang, T. Padır, A systematic review of robotics research in support of in-home care for older adults, *Information*, **11** (2020), 75. https://doi.org/10.3390/info11020075

12. W. Qi, A. Aliverti, A multimodal wearable system for continuous and real-time breathing pattern monitoring during daily activity, *IEEE J. Biomed. Health. Inf.*, **24** (2019), 2199–2207. https://doi.org/10.1109/JBHI.2019.2963048

13. C. Barras, Could speech recognition improve your meetings? *New Sci.*, **205** (2010), 18–19. https://doi.org/10.1016/S0262-4079(10)60347-8

14. Y. J. Lu, X. Chang, C. Li, W. Zhang, S. Cornell, Z. Ni, et al., Espnet-se++: Speech enhancement for robust speech recognition, translation, and understanding, preprint, arXiv:2207.09514.

15. L. Besacier, E. Barnard, A. Karpov, T. Schultz, Automatic speech recognition for under-resourced languages: a survey, *Speech Commun.*, **56** (2014), 85–100. https://doi.org/10.1016/j.specom.2013.07.008

16. G. I. Winata, S. Cahyawijaya, Z. Liu, Z. Lin, A. Madotto, P. Xu, et al., Learning fast adaptation on cross-accented speech recognition, preprint, arXiv:2003.01901.

17. S. Kim, B. Raj, I. Lane, Environmental noise embeddings for robust speech recognition, preprint, arXiv:1601.02553.

18. A. F. Daniele, M. Bansal, M. R. Walter, Navigational instruction generation as inverse reinforcement learning with neural machine translation, in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, (2017), 109–118. https://doi.org/10.1145/2909824.3020241

19. Z. Liu, D. Yang, Y. Wang, M. Lu, R. Li, Egnn: Graph structure learning based on evolutionary computation helps more in graph neural networks, *Appl. Soft Comput.*, **135** (2023), 110040. https://doi.org/10.1016/j.asoc.2023.110040

20. Y. Wang, Z. Liu, J. Xu, W. Yan, Heterogeneous network representation learning approach for Ethereum identity identification, *IEEE Trans. Comput. Social Syst.*, **10** (2022), 890–899. https://doi.org/10.1109/TCSS.2022.3164719

21. J. Zhao, Y. Lv, Output-feedback robust tracking control of uncertain systems via adaptive learning, *Int. J. Control Autom. Syst*, **21** (2023), 1108–1118. https://doi.org/10.1007/s12555-021-0882-6

22. S. Islam, A. Paul, B. S. Purkayastha, I. Hussain, Construction of English-bodo parallel text corpus for statistical machine translation, *Int. J. Nat. Lang. Comput.*, **7** (2018), 93–103. https://doi.org/10.5121/ijnlc.2018.7509

23. J. Su, J. Chen, H. Jiang, C. Zhou, H. Lin, Y. Ge, et al., Multi-modal neural machine translation with deep semantic interactions, *Inf. Sci.*, **554** (2021), 47–60. https://doi.org/10.1016/j.ins.2020.11.024

24. T. Duarte, R. Prikladnicki, F. Calefato, F. Lanubile, Speech recognition for voice-based machine translation, *IEEE Software*, **31** (2014), 26–31. https://doi.org/10.1109/MS.2014.14

25. D. M. E. M. Hussein, A survey on sentiment analysis challenges, *J. King Saud Univ. Eng. Sci.*, **30** (2018), 330–338. https://doi.org/10.1016/j.jksues.2016.04.002

26. Y. Liu, J. Lu, J. Yang, F. Mao, Sentiment analysis for e-commerce product reviews by deep learning model of bert-bigru-softmax, *Math. Biosci. Eng.*, **17** (2020), 7819–7837. https://doi.org/10.3934/mbe.2020398

27. H. Swapnarekha, J. Nayak, H. S. Behera, P. B. Dash, D. Pelusi, An optimistic firefly algorithm-based deep learning approach for sentiment analysis of COVID-19 tweets, *Math. Biosci. Eng.*, **20** (2023), 2382–2407. https://doi.org/10.3934/mbe.2023112

28. N. Mishra, M. Ramanathan, R. Satapathy, E. Cambria, N. Magnenat-Thalmann, Can a humanoid robot be part of the organizational workforce? a user study leveraging sentiment analysis, in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, (2019), 1–7. https://doi.org/10.1109/RO-MAN46459.2019.8956349

29. M. McShane, Natural language understanding (NLU, not NLP) in cognitive systems, *AI Mag.*, **38** (2017), 43–56. https://doi.org/10.1609/aimag.v38i4.2745

30. C. Li, W. Xing, Natural language generation using deep learning to support mooc learners, *Int. J. Artif. Intell. Educ.*, **31** (2021), 186–214. https://doi.org/10.1007/s40593-020-00235-x

31. H. Su, W. Qi, Y. Hu, H. R. Karimi, G. Ferrigno, E. De Momi, An incremental learning framework for human-like redundancy optimization of anthropomorphic manipulators, *IEEE Trans. Ind. Inf.*, **18** (2020), 1864–1872. https://doi.org/10.1109/tii.2020.3036693

32. Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, et al., Google's neural machine translation system: bridging the gap between human and machine translation, preprint, arXiv:1609.08144.

33. H. Hu, B. Liu, P. Zhang, Several models and applications for deep learning, in *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, (2017), 524–530. https://doi.org/10.1109/CompComm.2017.8322601

34. J. Aron, How innovative is apple's new voice assistant, Siri?, *New Sci.*, **212** (2011), 24. https://doi.org/10.1016/S0262-4079(11)62647-X

35. W. Jiao, W. Wang, J. Huang, X. Wang, Z. Tu, Is ChatGPT a good translator? Yes with GPT-4 as the engine, preprint, arXiv:2301.08745.

36. P. S. Mattas, ChatGPT: A study of AI language processing and its implications, *Int. J. Res. Publ. Rev.*, **4** (2023), 435–440. https://doi.org/10.55248/gengpi.2023.4218

37. H. Su, W. Qi, Y. Schmirander, S. E. Ovur, S. Cai, X. Xiong, A human activity-aware shared control solution for medical human–robot interaction, *Assem. Autom.*, **42** (2022), 388–394. https://doi.org/10.1108/AA-12-2021-0174

38. W. Qi, S. E. Ovur, Z. Li, A. Marzullo, R. Song, Multi-sensor guided hand gesture recognition for a teleoperated robot using a recurrent neural network, *IEEE Rob. Autom. Lett.*, **6** (2021), 6039–6045. https://doi.org/10.1109/LRA.2021.3089999

39. H. Su, A. Mariani, S. E. Ovur, A. Menciassi, G. Ferrigno, E. De Momi, Toward teaching by demonstration for robot-assisted minimally invasive surgery, *IEEE Trans. Autom. Sci. Eng.*, **18** (2021), 484–494. https://doi.org/10.1109/TASE.2020.3045655

40. J. Weizenbaum, ELIZA — a computer program for the study of natural language communication between man and machine, *Commun. ACM*, **26** (1983), 23–28. https://doi.org/10.1145/357980.357991

41. M. Prensky, Digital natives, digital immigrants part 2, do they really think differently? *Horizon*, **9** (2001), 1–6. https://doi.org/10.1108/10748120110424843

42. M. Skjuve, A. Følstad, K. I. Fostervold, P. B. Brandtzaeg, My chatbot companion-a study of human-chatbot relationships, *Int. J. Hum.-Comput. Stud.*, **149** (2021), 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

43. T. Kanda, T. Hirano, D. Eaton, H. Ishiguro, Interactive robots as social partners and peer tutors for children: a field trial, *Hum.-Comput. Interact.*, **19** (2004), 61–84. https://doi.org/10.1080/07370024.2004.9667340

44. J. Zakos, L. Capper, Clive-an artificially intelligent chat robot for conversational language practice, in *Artificial Intelligence: Theories, Models and Applications*, Springer, (2008), 437–442. https://doi.org/10.1007/978-3-540-87881-0_46

45. M. A. Salichs, Á. Castro-González, E. Salichs, E. Fernández-Rodicio, M. Maroto-Gómez, J. J. Gamboa-Montero, et al., Mini: A new social robot for the elderly, *Int. J. Social Rob.*, **12** (2020), 1231–1249. https://doi.org/10.1007/s12369-020-00687-0

46. J. Qi, X. Ding, W. Li, Z. Han, K. Xu, Fusing hand postures and speech recognition for tasks performed by an integrated leg–arm hexapod robot, *Appl. Sci.*, **10** (2020), 6995. https://doi.org/10.3390/app10196995

47. V. Lim, M. Rooksby, E. S. Cross, Social robots on a global stage: establishing a role for culture during human–robot interaction, *Int. J. Social Rob.*, **13** (2021), 1307–1333. https://doi.org/10.1007/s12369-020-00710-4

48. T. Belpaeme, P. Vogt, R. Van den Berghe, K. Bergmann, T. Göksun, M. De Haas, et al., Guidelines for designing social robots as second language tutors, *Int. J. Social Rob.*, **10** (2018), 325–341. https://doi.org/10.1007/s12369-018-0467-6

49. M. Hirschmanner, S. Gross, B. Krenn, F. Neubarth, M. Trapp, M. Vincze, Grounded word learning on a pepper robot, in *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, (2018), 351–352. https://doi.org/10.1145/3267851.3267903

50. H. Leeuwestein, M. Barking, H. Sodacı, O. Oudgenoeg-Paz, J. Verhagen, P. Vogt, et al., Teaching Turkish-Dutch kindergartners Dutch vocabulary with a social robot: does the robot's use of Turkish translations benefit children's Dutch vocabulary learning? *J. Comput. Assisted Learn.*, **37** (2021), 603–620. https://doi.org/10.1111/jcal.12510

51. S. Biswas, Prospective role of chat GPT in the military: according to ChatGPT, *Qeios*, 2023. https://doi.org/10.32388/8WYYOD

52. Y. Ye, H. You, J. Du, Improved trust in human-robot collaboration with ChatGPT, *IEEE Access*, **11** (2023), 55748–55754. https://doi.org/10.1109/ACCESS.2023.3282111

53. W. Qi, H. Su, A cybertwin based multimodal network for ECG patterns monitoring using deep learning, *IEEE Trans. Ind. Inf.*, **18** (2022), 6663–6670. https://doi.org/10.1109/TII.2022.3159583

54. W. Qi, H. Fan, H. R. Karimi, H. Su, An adaptive reinforcement learning-based multimodal data fusion framework for human–robot confrontation gaming, *Neural Networks*, **164** (2023), 489–496. https://doi.org/10.1016/j.neunet.2023.04.043

55. D. McColl, G. Nejat, Recognizing emotional body language displayed by a human-like social robot, *Int. J. Social Rob.*, **6** (2014), 261–280. https://doi.org/10.1007/s12369-013-0226-7

56. A. Hong, N. Lunscher, T. Hu, Y. Tsuboi, X. Zhang, S. F. dos R. Alves, et al., A multimodal emotional human–robot interaction architecture for social robots engaged in bidirectional communication, *IEEE Trans. Cybern.*, **51** (2020), 5954–5968. https://doi.org/10.1109/TCYB.2020.2974688

57. A. Meghdari, M. Alemi, M. Zakipour, S. A. Kashanian, Design and realization of a sign language educational humanoid robot, *J. Intell. Rob. Syst.*, **95** (2019), 3–17. https://doi.org/10.1007/s10846-018-0860-2

58. M. Atzeni, M. Atzori, Askco: A multi-language and extensible smart virtual assistant, in *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, (2019), 111–112. https://doi.org/10.1109/AIKE.2019.00028

59. A. Dahal, A. Khadka, B. Kharal, A. Shah, Effectiveness of native language for conversational bots, 2022. https://doi.org/10.21203/rs.3.rs-2183870/v2

60. R. Hasselvander, Buddy: Your family's companion robot, 2016.

61. T. Erić, S. Ivanović, S. Milivojša, M. Matić, N. Smiljković, Voice control for smart home automation: evaluation of approaches and possible architectures, in *2017 IEEE 7th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, (2017), 140–142. https://doi.org/10.1109/ICCE-Berlin.2017.8210613

62. S. Bajpai, D. Radha, Smart phone as a controlling device for smart home using speech recognition, in *2019 International Conference on Communication and Signal Processing (ICCSP)*, (2019), 0701–0705. https://doi.org/10.1109/ICCSP.2019.8697923

63. A. Ruslan, A. Jusoh, A. L. Asnawi, M. R. Othman, N. A. Razak, Development of multilanguage voice control for smart home with IoT, in *J. Phys.: Conf. Ser.*, **1921**, (2021), 012069. https://doi.org/10.1088/1742-6596/1921/1/012069

64. C. Soni, M. Saklani, G. Mokhariwale, A. Thorat, K. Shejul, Multi-language voice control iot home automation using google assistant and Raspberry Pi, in *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, (2022), 1–6. https://doi.org/10.1109/ACCAI53970.2022.9752606

65. S. Kalpana, S. Rajagopalan, R. Ranjith, R. Gomathi, Voice recognition based multi robot for blind people using lidar sensor, in *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*, (2020), 1–6. https://doi.org/10.1109/ICSCAN49426.2020.9262365

66. N. Harum, M. N. Izzati, N. A. Emran, N. Abdullah, N. A. Zakaria, E. Hamid, et al., A development of multi-language interactive device using artificial intelligence technology for visual impairment person, *Int. J. Interact. Mob. Technol.*, **15** (2021), 79–92. https://doi.org/10.3991/ijim.v15i19.24139

67. P. Vogt, R. van den Berghe, M. de Haas, L. Hoffman, J. Kanero, E. Mamus, et al., Second language tutoring using social robots: a large-scale study, in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, (2019), 497–505. https://doi.org/10.1109/HRI.2019.8673077

68. D. Leyzberg, A. Ramachandran, B. Scassellati, The effect of personalization in longer-term robot tutoring, *ACM Trans. Hum.-Rob. Interact.*, **7** (2018), 1–19. https://doi.org/10.1145/3283453

69. D. T. Tran, D. H. Truong, H. S. Le, J. H. Huh, Mobile robot: automatic speech recognition application for automation and STEM education, *Soft Comput.*, **27** (2023), 10789–10805. https://doi.org/10.1007/s00500-023-07824-7

70. T. Schlippe, J. Sawatzki, AI-based multilingual interactive exam preparation, in *Innovations in Learning and Technology for the Workplace and Higher Education*, Springer, (2022), 396–408. https://doi.org/10.1007/978-3-030-90677-1_38

71. T. Schodde, K. Bergmann, S. Kopp, Adaptive robot language tutoring based on bayesian knowledge tracing and predictive decision-making, in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, (2017), 128–136. https://doi.org/10.1145/2909824.3020222

72. B. He, M. Xia, X. Yu, P. Jian, H. Meng, Z. Chen, An educational robot system of visual question answering for preschoolers, in *2017 2nd International Conference on Robotics and Automation Engineering (ICRAE)*, (2017), 441–445. https://doi.org/10.1109/ICRAE.2017.8291426

73. C. Y. Lin, W. W. Shen, M. H. M. Tsai, J. M. Lin, W. K. Cheng, Implementation of an individual English oral training robot system, in *Innovative Technologies and Learning*, Springer, (2020), 40–49. https://doi.org/10.1007/978-3-030-63885-6_5

74. T. Halbach, T. Schulz, W. Leister, I. Solheim, Robot-enhanced language learning for children in Norwegian day-care centers, *Multimodal Technol. Interact.*, **5** (2021), 74. https://doi.org/10.3390/mti5120074

75. P. F. Sin, Z. W. Hong, M. H. M. Tsai, W. K. Cheng, H. C. Wang, J. M. Lin, Metmrs: a modular multi-robot system for English class, in *Innovative Technologies and Learning*, Springer, (2022), 157–166. https://doi.org/10.1007/978-3-031-15273-3_17

76. T. Jakonen, H. Jauni, Managing activity transitions in robot-mediated hybrid language classrooms, *Comput. Assisted Lang. Learn.*, (2022), 1–24. https://doi.org/10.1080/09588221.2022.2059518

77. F. Tanaka, T. Takahashi, S. Matsuzoe, N. Tazawa, M. Morita, Child-operated telepresence robot: a field trial connecting classrooms between Australia and Japan, in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (2013), 5896–5901. https://doi.org/10.1109/IROS.2013.6697211

78. A. S. Dhanjal, W. Singh, An automatic machine translation system for multi-lingual speech to Indian sign language, *Multimedia Tools Appl.*, **81** (2022), 4283–4321. https://doi.org/10.1007/s11042-021-11706-1

79. S. Yamamoto, J. Woo, W. H. Chin, K. Matsumura, N. Kubota, Interactive information support by robot partners based on informationally structured space, *J. Rob. Mechatron.*, **32** (2020), 236–243. https://doi.org/10.20965/jrm.2020.p0236

80. E. Tsardoulias, A. G. Thallas, A. L. Symeonidis, P. A. Mitkas, Improving multilingual interaction for consumer robots through signal enhancement in multichannel speech, *J. Audio Eng. Soc.*, **64** (2016), 514–524. https://doi.org/10.17743/jaes.2016.0022

81. Aldebaran, Thank you, gotthold! pepper robot boosts awareness for saving at bbbank, 2023. Available from: https://www.aldebaran.com/en/blog/news-trends/thank-gotthold-pepper-bbbank.

82. S. Yun, Y. J. Lee, S. H. Kim, Multilingual speech-to-speech translation system for mobile consumer devices, *IEEE Trans. Consum. Electron.*, **60** (2014), 508–516. https://doi.org/10.1109/TCE.2014.6937337

83. A. Romero-Garcés, L. V. Calderita, J. Martınez-Gómez, J. P. Bandera, R. Marfil, L. J. Manso, et al., The cognitive architecture of a robotic salesman, 2015. Available from: http://hdl.handle.net/10630/10767.

84. A. Hämäläinen, A. Teixeira, N. Almeida, H. Meinedo, T. Fegyó, M. S. Dias, Multilingual speech recognition for the elderly: the AALFred personal life assistant, *Procedia Comput. Sci.*, **67** (2015), 283–292. https://doi.org/10.1016/j.procs.2015.09.272

85. R. Xu, J. Cao, M. Wang, J. Chen, H. Zhou, Y. Zeng, et al., Xiaomingbot: a multilingual robot news reporter, preprint, arXiv:2007.08005.

86. M. Doumbouya, L. Einstein, C. Piech, Using radio archives for low-resource speech recognition: towards an intelligent virtual assistant for illiterate users, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **35** (2021), 14757–14765. https://doi.org/10.1609/aaai.v35i17.17733

87. P. Rajakumar, K. Suresh, M. Boobalan, M. Gokul, G. D. Kumar, R. Archana, IoT based voice assistant using Raspberry Pi and natural language processing, in *2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, (2022), 1–4. https://doi.org/10.1109/ICPECTS56089.2022.10046890

88. A. Di Nuovo, N. Wang, F. Broz, T. Belpaeme, R. Jones, A. Cangelosi, Experimental evaluation of a multi-modal user interface for a robotic service, in *Towards Autonomous Robotic Systems*, Springer, (2016), 87–98. https://doi.org/10.1007/978-3-319-40379-3_9

89. A. Di Nuovo, F. Broz, N. Wang, T. Belpaeme, A. Cangelosi, R. Jones, et al., The multi-modal interface of robot-era multi-robot services tailored for the elderly, *Intell. Serv. Rob.*, **11** (2018), 109–126. https://doi.org/10.1007/s11370-017-0237-6

90. L. Crisóstomo, N. F. Ferreira, V. Filipe, Robotics services at home support, *Int. J. Adv. Rob. Syst.*, **17** (2020). https://doi.org/10.1177/1729881420925018

91. I. Giorgi, C. Watson, C. Pratt, G. L. Masala, Designing robot verbal and nonverbal interactions in socially assistive domain for quality ageing in place, in *Human Centred Intelligent Systems*, Springer, (2021), 255–265. https://doi.org/10.1007/978-981-15-5784-2_21

92. S. K. Pramanik, Z. A. Onik, N. Anam, M. M. Ullah, A. Saiful, S. Sultana, A voice controlled robot for continuous patient assistance, in *2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec)*, (2016), 1–4. https://doi.org/10.1109/MEDITEC.2016.7835366

93. M. F. Ruzaij, S. Neubert, N. Stoll, K. Thurow, Hybrid voice controller for intelligent wheelchair and rehabilitation robot using voice recognition and embedded technologies, *J. Adv. Comput. Intell. Intell. Inf.*, **20** (2016), 615–622. https://doi.org/10.20965/jaciii.2016.p0615

94. A. Romero-Garcés, J. P. Bandera, R. Marfil, M. González-García, A. Bandera, Clara: Building a socially assistive robot to interact with elderly people, *Designs*, **6** (2022), 125. https://doi.org/10.3390/designs6060125

95. M. F. Ruzaij, S. Neubert, N. Stoll, K. Thurow, Multi-sensor robotic-wheelchair controller for handicap and quadriplegia patients using embedded technologies, in *2016 9th International Conference on Human System Interactions (HSI)*, (2016), 103–109. https://doi.org/10.1109/HSI.2016.7529616

96. T. Kobayashi, N. Yonaga, T. Imai, K. Arai, Bilingual SNS agency robot for person with disability, in *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, (2019), 74–75. https://doi.org/10.1109/GCCE46687.2019.9015297

97. C. Yvanoff-Frenchin, V. Ramos, T. Belabed, C. Valderrama, Edge computing robot interface for automatic elderly mental health care based on voice, *Electronics*, **9** (2020), 419. https://doi.org/10.3390/electronics9030419

98. D. Kottilingam, Emotional wellbeing assessment for elderly using multi-language robot interface, *J. Inf. Technol. Digital World*, **2** (2020), 1–10. https://doi.org/10.36548/jitdw.2020.1.001

99. Microsoft, What Is A Social Robot? 2021. Available from: https://codecondo.com/what-is-a-social-robot/.

100. Aldebaran, Pepper in the fight against COVID-19 at Horovice Hospital, Czech republic, 2023.

101. N. Shuo, S. Shao, N. Kubota, An iBeacon-based guide robot system for multi-lingual service, in *The Abstracts of the International Conference on Advanced Mechatronics: Toward Evolutionary Fusion of IT and Mechatronics: ICAM*, (2015), 274–275. https://doi.org/10.1299/jsmeicam.2015.6.274

102. S. Sun, T. Takeda, H. Koyama, N. Kubota, Smart device interlocked robot partners for information support systems in sightseeing guide, in *2016 Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (ISIS)*, (2016), 586–590. https://doi.org/10.1109/SCIS-ISIS.2016.0129

103. L. Jeanpierre, A. I. Mouaddib, L. Locchi, M. T. Lazaro, A. Pennisi, H. Sahli, et al., Coaches: an assistance multi-robot system in public areas, in *2017 European Conference on Mobile Robots (ECMR)*, (2017), 1–6. https://doi.org/10.1109/ECMR.2017.8098710

104. H. Yoshiuchi, T. Matsuda, J. Dai, Data analysis technology of service robot system for business improvement, in *ICRAI '19: Proceedings of the 5th International Conference on Robotics and Artificial Intelligence*, (2019), 7–11. https://doi.org/10.1145/3373724.3373733

105. A. Saniya, M. Chandana, M. S. Dennis, K. Pooja, D. Chaithanya, K. Rohith, et al., CAMPUS MITHRA: design and implementation of voice based attender robot, *J. Phys.: Conf. Ser.*, **2115** (2021), 012006. https://doi.org/10.1088/1742-6596/2115/1/012006

106. Q. Zhang, The application of audio control in social robotics, in *RICAI '22: Proceedings of the 2022 4th International Conference on Robotics, Intelligent Control and Artificial Intelligence*, (2022), 963–966. https://doi.org/10.1145/3584376.3584548

107. Aldebaran, Landscape AI: Robotic guides in museums and cultural places, 2023.

108. Y. Lin, H. Zhou, M. Chen, H. Min, Automatic sorting system for industrial robot with 3D visual perception and natural language interaction, *Meas. Control*, **52** (2019), 100–115. https://doi.org/10.1177/0020294018819552

109. B. Birch, C. Griffiths, A. Morgan, Environmental effects on reliability and accuracy of mfcc based voice recognition for industrial human-robot-interaction, *Proc. Inst. Mech. Eng., Part B: J. Eng. Manuf.*, **235** (2021), 1939–1948. https://doi.org/10.1177/09544054211014492

110. M. Kiruthiga, M. Divakar, V. Kumar, J. Martina, R. Kalpana, R. M. S. Kumar, Farmer's assistant using AI voice bot, in *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, (2021), 527–531. https://doi.org/10.1109/ICSPC51351.2021.9451760

111. J. H. Hong, J. Taylor, E. T. Matson, Natural multi-language interaction between firefighters and fire fighting robots, in *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, (2014), 183–189. https://doi.org/10.1109/WI-IAT.2014.166

112. J. Thomason, S. Zhang, R. J. Mooney, P. Stone, Learning to interpret natural language commands through human-robot dialog, in *IJCAI'15: Proceedings of the 24th International Conference on Artificial Intelligence*, (2015), 1923–1929. Available from: https://dl.acm.org/doi/10.5555/2832415.2832516.

113. R. Contreras, A. Ayala, F. Cruz, Unmanned aerial vehicle control through domain-based automatic speech recognition, *Computers*, **9** (2020), 75. https://doi.org/10.3390/computers9030075

114. Y. He, Z. Deng, J. Zhang, Design and voice-based control of a nasal endoscopic surgical robot, *CAAI Trans. Intell. Technol.*, **6** (2021), 123–131. https://doi.org/10.1049/cit2.12022

115. J. Nishihara, T. Nakamura, T. Nagai, Online algorithm for robots to learn object concepts and language model, *IEEE Trans. Cognit. Dev. Syst.*, **9** (2016), 255–268. https://doi.org/10.1109/TCDS.2016.2552579

116. H. M. He, Robotgpt: from Chatgpt to robot intelligence, 2023. https://doi.org/10.36227/techrxiv.22569247

117. F. Yuan, J. G. Anderson, T. H. Wyatt, R. P. Lopez, M. Crane, A. Montgomery, et al., Assessing the acceptability of a humanoid robot for Alzheimer's disease and related dementia care using an online survey, *Int. J. Social Rob.*, **14** (2022), 1223–1237. https://doi.org/10.1007/s12369-021-00862-x

118. S. Andrist, M. Ziadee, H. Boukaram, B. Mutlu, M. Sakr, Effects of culture on the credibility of robot speech: a comparison between English and Arabic, in *HRI '15: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, (2015), 157–164. https://doi.org/10.1145/2696454.2696464