International Journal of TESOL Studies (2025)
250908, 1-31 https://doi.org/10.58304/ijts.250908

Article

A Systematic Review on Artificial Intelligence (AI) Technologies in ESL/EFL Speaking Skills

Chuxuan Xing
Murad Abdu Saeed*
Universiti Malaya, Malaysia

Received: 2 August, 2025 / Received in revised form: 7 September, 2025 / Accepted: 7 September, 2025 / Available online: 11 September, 2025

Abstract
The rapid advancement of artificial intelligence (AI) technologies in language education has garnered significant attention from researchers who have explored AI applications in various language learning domains. Though there are numerous systematic reviews of studies on the application of AI in language education, speaking skills remain a crucial capacity for ESL/EFL learners to master. Most of these existing systematic reviews have primarily focused on areas such as language learning, teaching, and writing. Notably, only one systematic review has specifically addressed the use of AI Chatbots in speaking, leaving a gap in reviewing empirical research on this vital aspect of language learning. To bridge this research gap, this systematic review mainly investigated research trends on the applications, benefits, challenges, and implications of AI technologies in ESL/EFL speaking skills. It systematically reviewed 55 empirical studies about AI technologies in speaking skills published from 2017 to 2024, which were systematically selected from major academic databases (Web of Science, Scopus, ERIC, and Google Scholar) using the PRISMA flowchart and predefined inclusion/exclusion criteria. The data was generated from analysis of those empirical studies. The findings revealed that AI tools have several advantages in improving speaking skills, promoting learners' language proficiency and assisting students with pronunciation correction. However, the review also emphasized the potential limitations and challenges in using AI technologies in ESL/EFL speaking environments, such as inequality and accessibility, inaccurate speech recognition, and over-reliance. In addition, the review provided valuable technological, pedagogical, educational and research implications in the context of ESL/EFL speaking.

Keywords
Artificial intelligence (AI), ESL/EFL, AI in speaking skills, oral skills, systematic review

*Corresponding author. Email: muradsaeed@um.edu.my

1 Introduction

The rapid advancement of artificial intelligence (AI) technologies has profoundly impacted English language education, particularly in the field of speaking skills for English as a Second Language (ESL) and English as a Foreign Language (EFL) learners. AI made a difference in EFL learning, especially in automatic assessment, language translation, interaction between learners and teachers, and emotion management (Jiang, 2022). Speaking, regarded as a core but often most challenging language skill, it has been widely recognized as essential for ESL/EFL learners, however, it is also the skill where learners face the greatest obstacles due to classroom constraints and lack of exposure to real-life communication contexts (Aziz & Kashinathan, 2021). In recent years, AI applications have emerged as innovative tools to address persistent challenges in speaking instruction, such as limited authentic practice, lack of motivation, and insufficient communication opportunities (Imran et al., 2023). Efficient AI usage can boost learners' mastery of speaking skills and enhance language learners' motivation to speak in English.

However, existing systematic reviews on AI in language education often have a narrow focus and limited coverage of research topics. For example, Teng (2024) concentrates primarily on ChatGPT and provides little discussion of other AI writing tools. Other reviews focus mainly on writing, reading, and vocabulary skills, with less emphasis on the unique features and instructional complexity of speaking skills (Liang et al., 2023; Zhu & Wang, 2025). Moreover, most current reviews tend to stay at the level of theoretical discussion and tool classification, lacking critical analysis of teaching contexts, learner differences, and actual learning outcomes. For instance, while Wang et al. (2024a) explores different research paradigms from a philosophical perspective, it does not address deeper issues such as the homogeneity of AI tools, the marginalization of teachers and learners, or the short-term nature of learning gains. There is still a lack of comprehensive and authoritative empirical evidence on the actual impact and challenges of AI in ESL/EFL speaking instruction.

Distinct from previous reviews, this study systematically examines and critically analyzes empirical research on the use of AI in ESL/EFL speaking instruction published from 2017 to 2024. Unlike earlier work that often focused on a single type of tool or learner group, this review considers a wider range of AI technologies, learning contexts, and diverse learner populations. It also explores the deeper educational issues and complexities underlying AI applications in speaking instruction. By synthesizing the latest empirical evidence, this review aims to address theoretical and practical gaps in the field, and to offer forward-looking insights and recommendations for future research and teaching practice.

2 Literature Review

2.1 AI technologies in language learning

The application of AI in language learning has open up various new chances to improve efficiency and effectiveness, especially in speaking. Recent reviews have noted the growing trend of integrating AI into English speaking instruction, highlighting the versatility and accessibility of such tools (Hao et al., 2024). Intelligent chatbots, speech recognition apps, automated feedback tools, and mobile-assisted language learning (MALL) platforms enable personalized, interactive learning tailored to individual needs (Fathi et al., 2024; Rajendran & Yunus, 2021). For example, platforms such as voice-interactive AI chatbots, Automatic Speech Recognition (ASR) and ELSA provide learners with personalized feedback and unique learning experience, helping learners engage in self-regulating behaviors and supporting the cyclic learning process through personalized learning partners(Chang & Sun, 2024; Koç & Savaş, 2024; Pham & Pham, 2025). Mobile learning solutions and AI avatars help create engaging, learner-centered environments, reduce learning anxiety and increase learning interest and willingness to communicate (Kang 2022; Rajendran & Yunus, 2021). Information and communication technology (ICT)

and AI speech assessment programs could significantly enhance language learners' speaking fluency, pronunciation, and expression, with notable improvements in overall proficiency and vocabulary. (Madhavi et al., 2023). In recent years, the rapid advancement of generative AI, particularly following the emergence of large language models (LLM) like GPT-3 in 2020 and the launch of ChatGPT in late 2022, has begun to transform the landscape of language education. Unlike earlier AI tools, generative AI can engage in extended conversations, simulate real-world scenarios, and provide context-sensitive feedback, giving learners access to more authentic and interactive speaking practice (Toncelli & Kostka, 2024; Ulla & Teng, 2024). These developments have brought new opportunities for personalized and contextualized instruction in speaking classes, helping students build confidence, develop critical thinking, and foster greater autonomy (Huang, 2024; Yamaoka, 2024).

2.2 Empirical evidence: Benefits of AI for speaking

Recent empirical studies have demonstrated the positive effects of AI technologies on ESL/EFL learners' speaking skills. Fathi et al. (2024) conducted a rigorous experimental study with 65 Iranian intermediate EFL learners (aged 18–25), who were randomly assigned to an AI-mediated group (n = 33) or a traditional face-to-face group (n = 32). Over 12 weeks, the AI group interacted daily with Andy English Chatbot for at least 20 minutes, while both groups also participated in identical classroom speaking activities. Speaking proficiency was measured using the IELTS speaking test (covering fluency, lexicon, accuracy, and pronunciation) and a willingness to communicate (WTC) scale. Results showed that the AI group made significantly greater gains than the face-to-face group across all aspects of speaking skills and WTC, after controlling pretest differences. Specifically, the AI group's mean total speaking score rose from 5.12 to 6.67, compared to an increase from 4.89 to 5.89 in the control group. ANCOVA results showed a significant and large effect for the AI intervention in improving learners' speaking skills and willingness to communicate. Qualitative interviews indicated that the chatbot's personalized feedback, low-pressure environment, and instant correction boosted learners' confidence and motivation, although some still preferred human interaction for more detailed feedback and social engagement.

In addition to qualitative findings, qualitative research has provided deeper insight into AI's impact on learner motivation and classroom dynamics. Yamaoka (2024) conducted open-ended questionnaire with ten Japanese university EFL students, all the participants had no prior experience with ChatGPT. The students used ChatGPT primarily for translation, learning expressions, summarizing, and proofreading. Most participants viewed ChatGPT as an effective learning supporter, highlighting that its immediate assistance reduced anxiety and provided reassurance; for example, one student said they felt "more reassured when receiving advice from ChatGPT than when thinking alone." However, some students, particularly those with higher English proficiency, reported decreased self-confidence when comparing their writing to ChatGPT's output, and some were concerned about the reliability of AI feedback. Yamaoka emphasized the importance of teacher guidance and digital literacy training to help students use ChatGPT effectively and avoid negative effects.

Building on these findings, Huang (2024) explored ChatGPT's voice feature for speaking feedback by designing tailored prompts for pronunciation, structure, grammar, and coherence. In classroom field tests, students practiced oral presentations and pronunciation with ChatGPT, receiving immediate, constructive feedback. Observations showed that this approach created a supportive, low-pressure environment, helping students reduce anxiety and participate more actively. Teachers also saw value in using the tool for speech rehearsal, especially for less confident learners. However, the study noted that AI feedback cannot fully substitute for real human interaction, particularly regarding delivery skills, audience engagement, and authentic communicative competence. There is a need for ongoing teacher involvement, careful prompt design, and digital literacy training so students can use AI-generated feedback effectively and avoid over-reliance. These studies show that tools like chatbots and generative

models can help EFL learners improve their speaking skills, motivation, and classroom participation. However, their benefits depend on proper guidance from teachers and effective use in real lessons.

These studies demonstrate that AI-based tools such as chatbots and generative models can significantly enhance EFL learners' speaking skills, motivation, and classroom participation. However, to get the most out of these tools, teachers need to give proper guidance and use them well in real lessons. It is also important to teach students how to use these tools wisely and for teachers to stay involved, so students can benefit and avoid possible problems.

2.3 Challenges and limitations of AI integration

Despite these benefits, several challenges complicate the effective use of AI technologies in speaking instruction. Problems related to technology, interaction, skill development and academic standards mean that AI must be thoughtfully guided and integrated to truly help students improve their language abilities (Teng, 2024). One main issue is that AI often struggles with speech and accent recognition, so the feedback it provides is not always accurate or suitable for learners with different backgrounds (Zou et al., 2020). This problem is especially obvious in correcting surface-level errors in speaking and writing (Teng, 2024). In addition, AI systems lack the ability to fully understand complex situations and real-life communication. As Floridi and Chiriatti (2020) observe, AI-generated responses are based on pattern matching rather than true understanding of meaning or context. Crompton and Burke (2023) also point out that most AI systems in higher education are still mainly used for routine or well-defined tasks. They note that current research and practice have not yet addressed how AI could handle more complex or interactive communication in real educational contexts. These issues make it harder for learners to build real communication skills and to interact effectively across cultures.

Importantly, overreliance on AI may undermine learners' critical thinking, creativity, and independent expression. In Sherwood & Mac Donald (2024), some participants explicitly voiced concerns that excessive dependence on AI tools like ChatGPT could diminish their critical thinking abilities, as reflected in the coding category 'Reduction of critical thinking' (5.81% of negative responses) and in student comments describing the tendency to bypass critical analysis and reflection during the writing process. The authors highlight the urgent need to embed critical thinking activities into AI-assisted instructions to mitigate these risks. These findings echo broader scholarly concerns about responsible integration of AI in education. As Barrot (2023) and Teng (2024) emphasize, it is essential to promote balanced and reflective use of AI tools to ensure students maintain active engagement and critical participation. Additionally, the rise of AI-generated content has intensified challenges related to academic integrity and review reliability. Risks such as fabricated references, inaccurate information, and blurred boundaries between human and machine writing complicate the task of maintaining academic standards and quality assurance (Teng, 2023). Careful oversight and explicit pedagogical guidelines are necessary to address these evolving issues.

2.4 Current gaps in literature

Although several systematic reviews have addressed AI's role in language learning, most of these investigated the use of AI in second language acquisition (SLA) teaching and Applied linguistics (Kartal & Yeşilyurt, 2024), self-regulated language learning (Chang & Sun, 2024) and learning from self-determined theory (Li, Zhou, & Chiu, 2025) as well as emotional AI in English language education (Liu et al., 2024). In addition, Koç and Savaş (2024) systematically reviewed several 56 studies about AI Chatbots used in language learning which were published from 2010 to 2024, identifying three types of AI Chatbots. Yet, this systematic review was exclusive to theoretical frameworks, methodological and technological properties, learners' experience and pedagogical implementations. Moreover, another

systematic review by Wang et al. (2024b) reviewed 49 studies (2008-2022) about human-AI interaction. The results showed that speaking ranked third in the language domains where learners interact with AI, behind overall skills and vocabulary. Notably, learners who interacted with the AI achieved better verbal scores, including speaking, than those who did not. Du and Danie (2024) carried out a systematic review of 24 studies regarding the usage of AI-chatbots in English language education. Research has found that AI chatbots accelerated English learning by supporting spoken language goals, reducing anxiety, and increasing engagement and confidence. However, it only focused on chatbots, excluding tools such as ChatGPT, and only covered research through 2023.

2.5 Purpose of the present review and research questions

Research in the field of AI used in language education has grown significantly since 2017. Zawacki-Richter et al. (2019) noted that the topics of AI-related studies in higher education have continued to expand, encompassing applications such as intelligent tutoring and automated assessment. More recently, Zhu and Wang (2025) identified that advances in technologies like deep learning have driven the emergence of new research directions, including automated feedback, writing evaluation, and intelligent conversational agents, which have contributed to a rapid increase in the volume of related studies, this study will review these empirical studies published from 2017 to 2024 systematically to construct a comprehensive understanding of AI in EFL/ESL speaking skills from multiple perspectives, with the intention of highlighting the present limitations and challenges of using AI to support speaking skills, the research will provide key directions and advice for promotion of future research. Specifically, the current systematic review addressed the following research questions:

1. What AI technologies are used in ESL/EFL speaking contexts?
2. What are the contextual and methodological contexts of the selected empirical studies on AI in ESL/EFL speaking skills?
3. In what ways do ESL/EFL students benefit from AI tools in the ESL/EFL speaking contexts? 4. What are the potential limitations and challenges involved in integrating AI technologies in ESL/EFL speaking contexts?

3 Method

This research aimed to systematically review the application of AI technologies in EFL/ESL language learners' speaking skills. A systematic review uses clear, effective methods to reduce bias, provide more reliable findings, and draw valid conclusions (Munn et al., 2018). Applying this method, this research will provide strong insights into the impact of AI technologies in improving speaking skills.

3.1 Inclusion and exclusion criteria

Before conducting this systematic review, we followed a rigorous and systematic approach to guarantee the reliability and relevance of papers selected. The process was guided by clearly defined inclusion and exclusion criteria. First, only articles published in English were included; studies in other languages were excluded. Second, this review focused on the use of AI technologies to improve the English-speaking skills of ESL/EFL learners, so studies mainly about writing, listening, vocabulary, or other language skills were excluded. Third, only empirical studies with clear methods, data collection, and analysis were included. Non-empirical studies, such as conceptual papers, review articles, or those focused only on the design or theory of AI systems, were excluded because they did not match the practical and evidence-based focus of this review. Fourth, only studies published in peer-reviewed journals were

considered, to ensure academic quality. Conference papers, book chapters, and other informal sources were not included. To reflect the latest developments in the field, only studies published between 2017 and 2024 were included. At the time of our literature search and analysis, most 2025 articles had not yet been published or indexed in major academic databases. Using 2024 as the cut-off year ensured that our review was based on a complete and accessible set of studies and allowed the screening process to be rigorous and reproducible. Articles published outside this period were excluded to ensure that the review remains up to date.

In practice, all records were independently screened by two reviewers at the title, abstract, and full-text stages. Any disagreements were resolved through discussion until consensus was reached. For each study excluded after full-text screening, the specific reason for exclusion was carefully documented. A complete list of excluded studies and their reasons. By following this rigorous and transparent process, we ensured that all included studies were highly relevant to the research topic and met high methodological and academic standards, providing a solid foundation for this systematic review.

3.2 Article searching and selecting processes

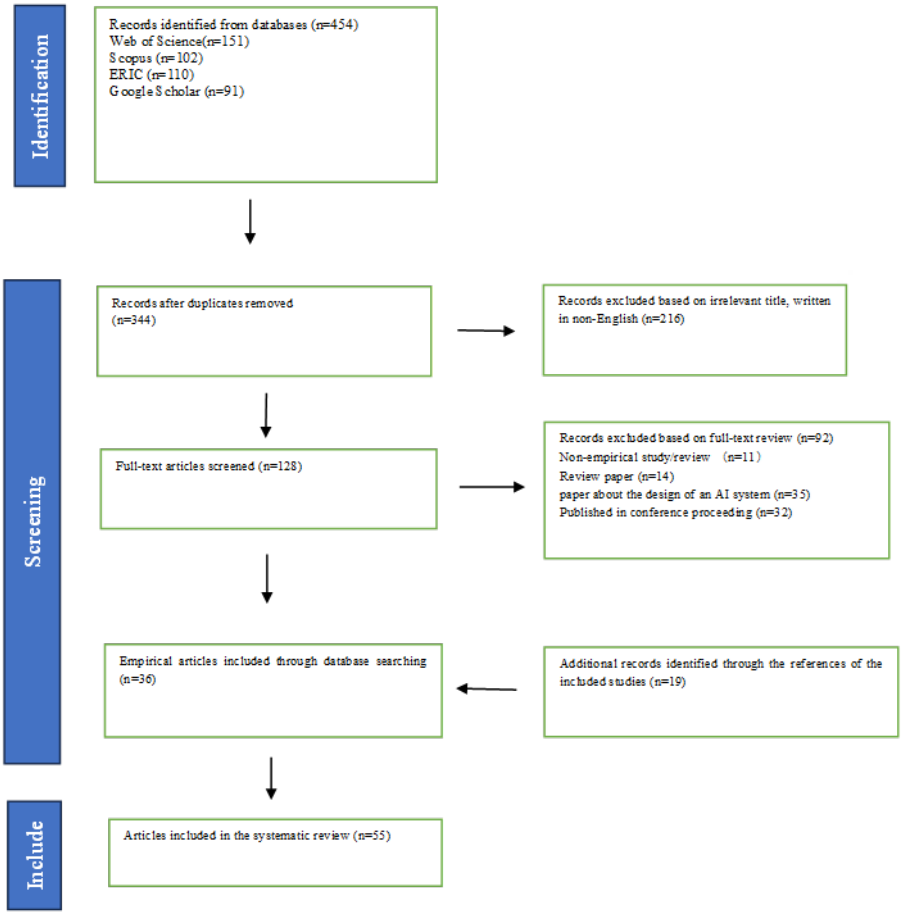
This systematic review employed a structured approach, utilizing meta-analysis (PRISMA) (Figure 1) to analyze the included peer-reviewed papers and ensure a rigorous and transparent review process. To retrieve relevant literature, we searched for the Web of Science (WoS) (n=151), Scopus (n=102), ERIC (110), Google Scholar (n=91), and the homepages of relevant journals (e.g., Computer Assisted Language Learning, System and Sustainability). Web of Science and Scopus was selected for their comprehensive coverage of high-quality, peer-reviewed journals. ResearchGate was used to access scholarly publications and connect directly with researchers, while Google Scholar was included to capture additional relevant literature that might not be indexed in other databases. The search strategy incorporated several keyword groups, such as “AI in Language Learning,” “AI for Speaking,” “Chatbot for Speaking,” “ChatGPT for Speaking,” “Chatbot for Speaking English,” and “ChatGPT for Speaking English.” The keywords were combined using Boolean operators to enhance the comprehensiveness and relevance of the search (Scells et al., 2020). Specifically, “OR” was used within each group to include synonyms and related terms, while “AND” was used between groups to ensure that retrieved studies addressed all relevant aspects. For example, a typical search string was: (“AI” OR “Chatbot” OR “ChatGPT”) AND (“Language Learning” OR “Speaking” OR “Speaking English”). This strategy was applied consistently across all databases.

The data collection process began on November 2, 2024, and lasted for three weeks. In the initial search, 454 articles were identified. After retrieving all records from the selected databases, we imported them into EndNote for management. Duplicate articles were first identified and removed using EndNote’s built-in duplicate detection tool. Following this automated procedure, we conducted a manual review to ensure that all remaining duplicates were eliminated. As a result, 344 unique studies were retained and subsequently subjected to a three-phase screening process. In the first phrase, Titles and abstracts were screened to determine relevance, resulting in the exclusion of 216 articles that were either irrelevant or not written in English. The remaining 128 articles underwent a full-text review. At this stage, 92 articles were excluded for not meeting the inclusion criteria, such as being non-empirical studies, reviewing articles, or focusing primarily on the design of AI systems. As a result, 36 empirical studies were included based on database searches. Additionally, 19 more articles were identified through a review of the reference lists of the included studies. In total, 55 articles were included in the final analysis, providing a solid foundation for this systematic review.

The methodological quality of the included studies was assessed using the Mixed Methods Appraisal Tool (Hong et al., 2018). This tool is designed for evaluating qualitative, quantitative, and mixed methods

studies. Each study was independently evaluated by two reviewers according to the MMAT criteria. Disagreements were resolved through discussion.

Figure 1
PRISMA Flowchart of Article Selection



Alt Text: A PRISMA flowchart illustrating the screening and selection process of studies included in the systematic review. It shows the number of records identified, screened, excluded, assessed for eligibility, and ultimately included in the final synthesis.

3.3 Data analysis and coding of selected studies

To analyze and code the selected studies systematically, contextual information from eleven aspects were synthesized (see appendix A): year of the publication, AI technology, study design, simple size, study setting, learner type, context/country, data collection and analysis, potential/benefits, limitations/challenges, implications. The former seven aspects were coded based on the description of the text, and the later four aspects were coded using thematic analysis (Thomas & Harden, 2008). Below is the interpretation about how each question was addressed, the tools and frameworks applied, and the rationale behind the coding and explanation.

The "author/year" column is recorded in the format of "first author's surname and publication year," serving as a unique identifier for each study. To clearly understand the main purpose and role of these AI tools used in language learning, these tools are then categorized according to their core technological functions and application context. Methods outlined in the "Study Design" section were also coded as quantitative, qualitative, and mixed, providing insights into research rigor and design trends. The sample

size and learner type further enriched our understanding of the context, allowing for a nuanced analysis of how different settings and approaches affect research outcomes. Contextual and methodological details were systematically extracted from the data analysis, learning settings, "formal learning" and "informal learning", were recorded and classified to reflect the different contexts in which AI tools are studied. "Learner type" is primarily coded based on participants' English proficiency levels or learning stages (e.g., A1–A2, B1–B2), revealing the coverage of different learner groups. "Context/country" indicates the geographical and cultural background of the study, specified by country or region. To clearly identify how each study interprets its data and to allow for meaningful comparison of research approaches and result reliability, "Data collection and analysis" section codes the specific data collection tools used in each study, such as questionnaire, interview.

The benefits were determined by reviewing the results described in the "Potential/Benefits" section. Through thematic analysis, four main themes were identified: "improving language skills", "improving speaking confidence", "personalized and adaptive learning", and "flexibility and accessibility". Further analysis linked these advantages to the specific functions of AI tools. Among them, "improving language skills" covered "improving speaking fluency and pronunciation"; "improving speaking confidence" included "enhancing speaking participation", "stimulating learning motivation", and "alleviating speaking anxiety"; "personalized and adaptive learning" focused on "providing personalized learning experience and real-time feedback". This classification clearly showed the multiple advantages of AI tools in supporting language learners.

To analyze obstacles in using AI for language learning, we coded them as "limitations" and "challenges" to distinguish between the inherent weaknesses or constraints of AI technologies ("limitations") and the practical or contextual difficulties encountered during implementation ("challenges"). Using thematic analysis, it is possible to classify the themes of limitations and challenges into technical, cultural, pedagogical, and ethical areas. Technical challenges include issues such as inequalities in access to AI technologies, limitations in the accuracy of speech recognition, and lack of social interaction and emotional engagement. The latter consists of problems related to contextual and cultural sensitivity. From a pedagogical perspective, a major concern is overdependence on AI technologies and a lack of social interaction. Lastly, the main ethical challenge is in relation to privacy issues. Such categorization brought clarity of structure for understanding the multidimensionality in the limitations presented by the combination of AI into the ESL/EFL speaking context.

Finally, to ensure inter-rater reliability, two coders were involved in coding and analyzing the literature review. Both coders independently analyzed the data using Microsoft Excel. Initially, the inter-coder agreement rate was 83%. Through many rounds of discussion, most discrepancies were resolved, and the final agreement rate reached 87.5%, which is generally considered acceptable in previous studies (Krippendorff, 2013). Any remaining disagreements were resolved through consensus.

4 Findings

4.1 What AI technologies are used in ESL/EFL speaking contexts?

Figure 2 and Figure 3 indicated that AI chatbots have become the dominant force, accounting for nearly half of all applications and showing a continuous rise in usage frequency in recent years. Chatbots are not merely responsive tools; they have become vital platforms for learners' daily speaking input, interactive output, and autonomous expression. By leveraging context, chatbots can simulate a wide range of authentic communicative scenarios, lower learners' anxiety about speaking, and offer sustained feedback and emotional support. In some cases, they even replace traditional interpersonal speaking practice. For example, Fathi et al. (2024) found that Andy English Chatbot creates a low-pressure environment, reducing anxiety caused by peer evaluation and encouraging learners to speak more actively, thereby enhancing confidence and willingness to communicate.

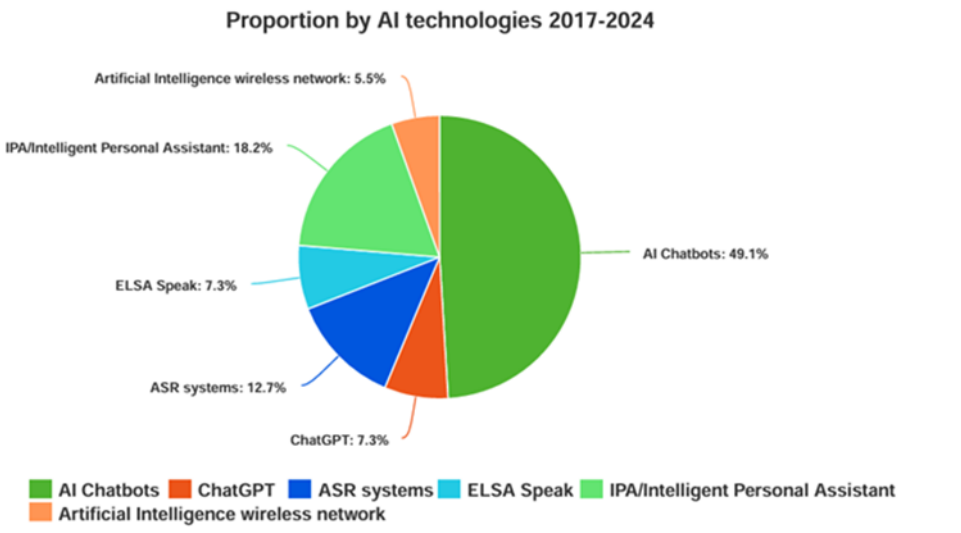
Nevertheless, the use of AI in speaking instruction extends beyond chatbots. Intelligent personal assistants (IPAs) such as Alexa (Dizon, 2020) and Google Assistant (Kim et al., 2021) have also seen increasing adoption in recent years. Their core advantage lies in providing learners with opportunities for spoken interaction that closely resemble real-life scenarios, thus extending language learning beyond classrooms and textbooks into daily contexts. Interaction with IPAs allows learners to practice pragmatic competence and improve responsiveness in authentic communication, a sense of realism and flexibility that is difficult to achieve in traditional instruction. At the same time, the widespread adoption of automatic speech recognition (ASR) has further improved oral assessment systems. ASR can instantly transcribe spoken input and, when integrated with intelligent scoring systems, quantitatively assess multiple dimensions such as pronunciation, intonation, and fluency. This enables both teachers and learners to obtain more objective and detailed feedback, thus enhancing the scientific and targeted nature of oral training. Applications like Duolingo, which are based on ASR, have been widely used in language learning and have shown significant positive effects on speaking and listening skills, learner autonomy, motivation, and flexibility, making them effective tools for improving communicative competence in English (Azzahra et al., 2024; Qiao & Zhao, 2023).

More recently, the emergence of large generative language models such as ChatGPT has opened new possibilities for speaking language learning. Although their overall proportion remains lower than that of other tools, heatmap data show a marked increase in usage after 2023. Compared with traditional chatbots, large language models possess a stronger capacity for contextual understanding and response generation, and can dynamically adjust topics, styles, and depth of interaction according to learners' needs. This advanced conversational capability provides a technical foundation for creating immersive and personalized learning environments and greatly expands the potential of AI-driven speaking instruction.

At the same time, specialized AI tools like ELSA Speak continue to play a key role in foundational pronunciation training and self-correction. While their total usage is not as high as that of chatbots, such tools are indispensable for raising learners' phonological awareness and overcoming negative transfer from their first language. It is also worth noting that some AI technologies, such as AI-powered wireless networks, though accounting for a smaller share of statistics, provide essential support for remote speaking ion and the realization of intelligent classrooms. These technologies optimize data transmission and real-time interaction, ensuring the seamless integration of AI tools into online and blended learning environments.

Figure 2

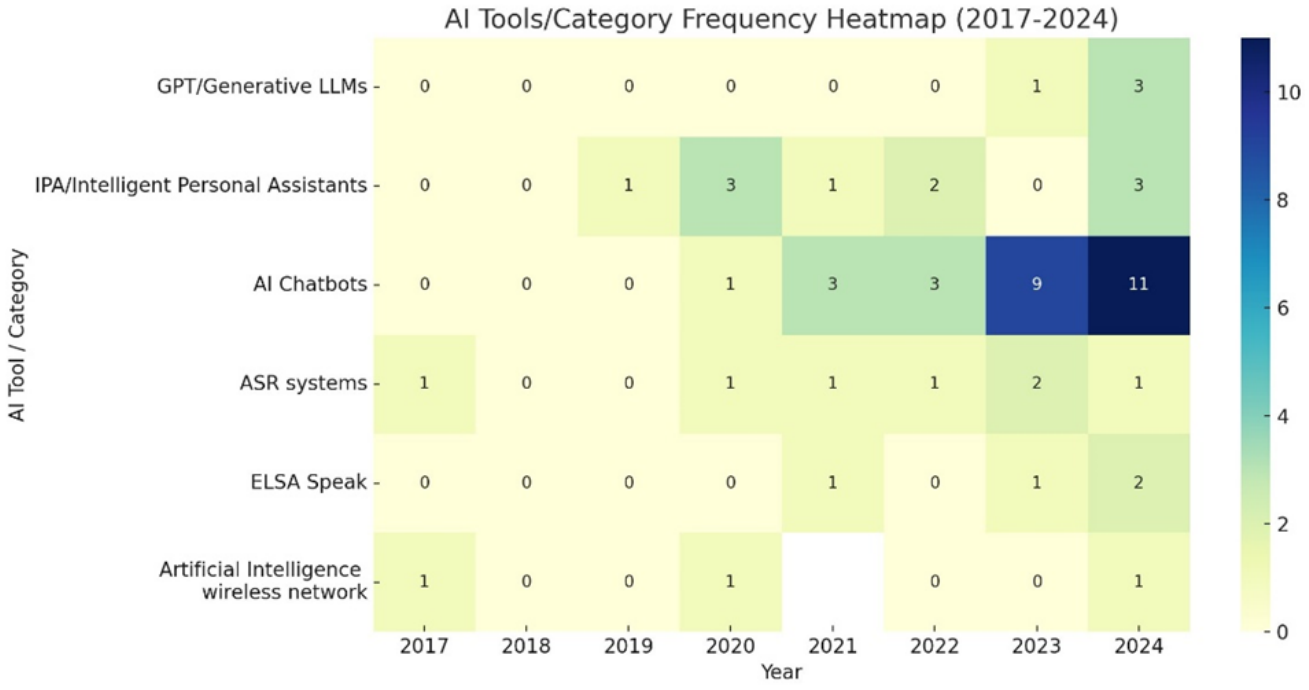
Proportion of AI Technologies Used in Language Learning Studies (2017–2024)



Alt Text: Pie chart displaying the proportion of different AI technologies applied in language learning studies from 2017 to 2024.

Figure 3

AI Tools/Category Frequency Heatmap (2017–2024)



Alt Text: Heatmap showing the yearly use frequency of various AI tool categories in language learning from 2017 to 2024

4.2 What are the contextual and methodological contexts of the selected empirical studies on AI in ESL/EFL speaking skills?

This paper reviewed 55 empirical studies on AI in ESL/EFL speaking Settings, highlighting different learning environments and methodological approaches. As for the year of publication (Figure 4), these selected studies were published between 2017 to 2024, and the research on AI tools used in EFL/ESL speaking showed a significant increase, especially after 2020, indicating a great research interest in this field. In terms of contextual context, a significant proportion of the research focused on formal learning (n=53), such as in universities and language institutions, while a smaller subset explored informal learning (n=2). Figure 5 clearly demonstrated a concentration of research on learners at the B1–B2 (intermediate) level, with this group comprising most of the sample (n=30). Although there is a relatively substantial representation of learners at the A1–A2 (foundation) level (n=14), their numbers are still considerably lower than those of the intermediate group. Learners at the A2–B1 (lower intermediate) stage are less represented (n=8), while mixed and extended proficiency groups such as A1–B2, A2–B2, and B1–C1 are scarcely present, each accounting for just one study. This distribution highlighted a notable imbalance in participant selection within the existing research on technology-enhanced speaking instruction. The predominance of intermediate-level learners may be attributed to their greater capacity for autonomous learning and engagement with complex instructional tasks.

Figure 4

A Bar Chart Showing the Publication of 55 Selected Articles

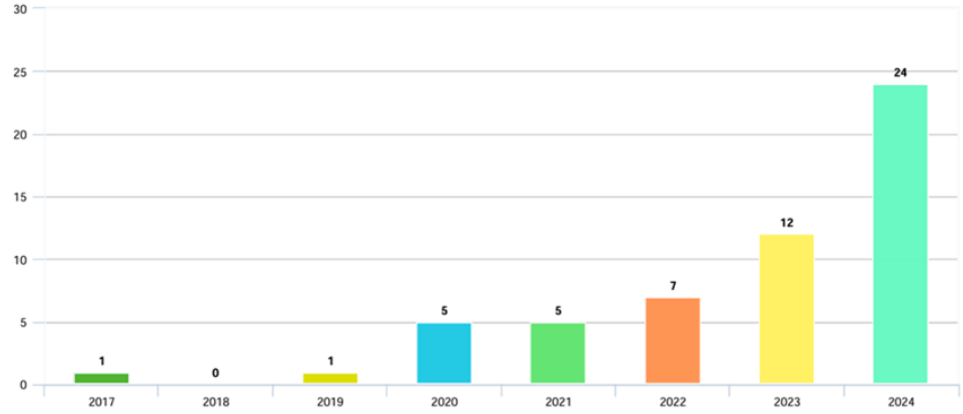
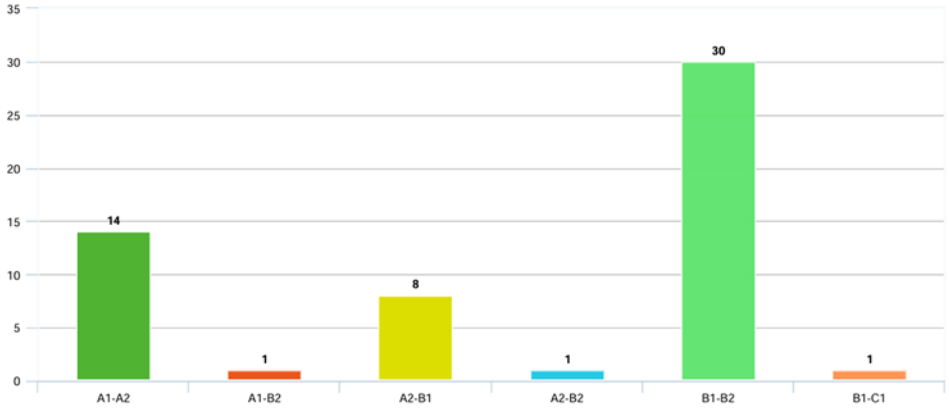


Figure 5

A Bar Chart Showing Learner Type

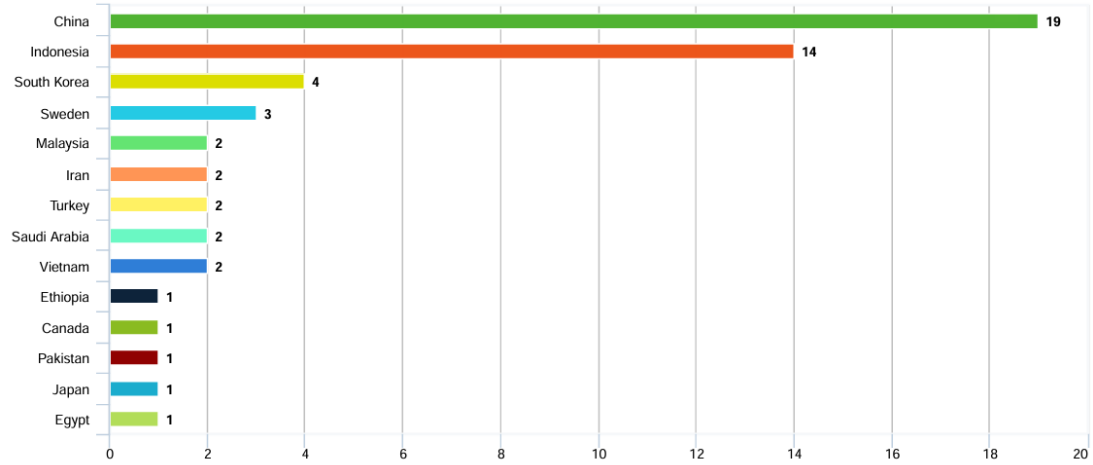


Alt Text: The chart shows the highest concentration of studies involved B2–C1 (n=37) college-level learners, followed by A1–A2 (n=14), A1–B1 (n=8), A2-B2 (n=1), A1-B2 (n=1), B1-B2 (n=30) and B1-C1 (n=1). The chart highlights limited focus on beginner and foundation-level learners.

Geographically, these studies span a wide range of regions, reflecting the global interest in leveraging AI to enhance spoken English proficiency among non-native speakers. As shown in Figure 6, China contributed the highest number of studies (n=19), followed by Indonesia (n=14), South Korea (n=4) and Sweden (n=3). Other countries such as Sweden (n=3), Malaysia, Iran, Saudi Arabia, Turkey and Vietnam (each n=2) were also represented. In contrast, several countries, including Ethiopia, Canada, Pakistan, Japan and Egypt were each the focus of only one study (n=1), indicating a relatively limited research presence in these contexts. This concentration can largely be attributed to the high priority given to AI-driven language education in some Asian countries, as well as the continuous support from policies and investment in resources. The dominant position of China and Indonesia in related research reflected the two countries' proactive efforts in educational reform and technology integration. In contrast, research in this field remains relatively limited in traditional educational powers such as Europe and North America, highlighting regional imbalances and gaps in the global application of AI for spoken English instruction. Such a research focus not only restricts the global applicability of AI-based oral English teaching outcomes but also risks overlooking the impact of local needs and cultural differences on the effectiveness of technology implementation.

Figure 6

A Bar Chart Showing Countries Where These Studies Were Conducted

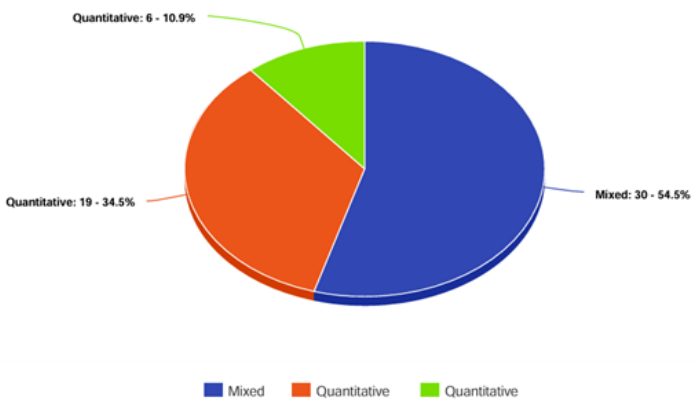


Alt Text: A bar chart showing the number of reviewed studies by country. Countries that reflect global but uneven adoption of AI in ESL/EFL speaking research.

Methodologically, as illustrated in Figure 7, most of the included studies adopted a mixed-methods approach (n=30), accounting for 54.5%, while studies relying solely on quantitative (n=19) and qualitative methods (n=6) represented 34.5% and 10.9% respectively. The predominance of mixed methods suggested that researchers in this field increasingly recognize the complexity of technology-supported language learning and thus favor combining quantitative measures, such as pre- and post-test designs and statistical analyses of achievement, with qualitative insights from interviews, learner reflections, and thematic analysis. For instance, Zhang et al. (2024) employed a mixed-methods design and found that chatbots significantly enhanced Chinese English learners' willingness to communicate, whereas intelligent tutoring systems (ITS) were less effective. Purely quantitative studies (n=19) tended to focus on the direct impact of technological tools on speaking proficiency, while qualitative research (n=6) provided deeper understanding of learners' experiences and interactional dynamics. For example, Azzahra et al. (2024) found that English education students generally view Duolingo and ELSA Speak as helpful for improving listening and speaking skills, but note issues like limited free features, reliance on stable internet, and accent recognition challenges. This widespread use of mixed methods reflected the value placed on integrating diverse perspectives to achieve a more comprehensive understanding of language learning outcomes. Nevertheless, the relatively small number of purely qualitative studies indicates a need for further exploration of learners' subjective experiences and contextual differences, which are often beyond the reach of quantitative measures.

Figure 7

A Pie Chart Showing Research Method Distribution



Alt Text: A bar chart illustrating the distribution of research methods used in the reviewed studies. Mixed methods dominate (n=31, 54.5%), followed by quantitative (n=19, 34.5%) and qualitative approaches (n=8, 10.9%), highlighting a strong preference for methodological integration in AI-related ESL/EFL speaking research.

These studies varied in sample size, but were small overall, ranging from 6 (Azzahra et al. 2024) to 366 people (Zou et al., 2024), with most studies involving fewer than 100 participants. The limited size of this sample may restrict the applicability of the findings to a broader population. The duration of the studies was usually short. This raises concerns about the lasting sustainability of AI integration. Though these studies provide valuable implications with the immediate effect of AI on the advancement of spoken language skills, they often fail to assess the long-term effects, such as retention of language skills or changes in learner autonomy.

4.3 In what ways do ESL/EFL students benefit from AI tools in the ESL/EFL speaking contexts?

By conducting a thematic analysis of the data, we identified several key themes (Table 1) that reflected the ways in which ESL/EFL students benefit from AI tools in speaking contexts. These themes capture the core experiences and advantages students reported in their use of AI tools.

Table 1
Benefits of AI for Speaking

Themes	Number of Studies on That Theme	Percentage (%)
Improved speaking fluency	45	82%
Improved pronunciation	29	53%
Increased confidence in speaking	17	31%
Personalized learning	16	29%
Real-Time and real-time feedback	11	20%
Increased engagement and motivation	21	38%
Flexibility	5	9%
Accessibility	4	7%
Reduce speaking anxiety	19	35%

4.3.1 Improved speaking fluency and pronunciation

AI tools have significantly enhanced the speaking fluency and pronunciation of language learners. Many participants reported that frequent practice with AI-driven tools, such as Duolingo and ELSA Speak, improved their speaking skills. These tools offer real-time feedback, enabling learners to quickly identify and correct pronunciation errors, thereby accelerating language acquisition, particularly in terms of accent and fluency. Saraswati et al. (2023) noted that interacting with AI-powered robots like Replika helped learners correct pronunciation mistakes, expand their vocabulary, and boost their speaking confidence. Similarly, research by Annamalai et al. (2023) highlighted the strong performance of AI tools in correcting pronunciation and listening errors, leading to improvements in learners' English proficiency. In addition, the personalized learning features of AI tools provide tailored feedback based on the specific needs of learners, which helps enhance their confidence and language proficiency. Practice targeting specific pronunciation issues has led to significant progress in learners' language abilities (Khailifa & Ginting, 2024; Kholis, 2021; Moussalli & Cardoso, 2020).

4.3.2 Increased engagement and motivation

The interactive and gamified features of AI tools such as Duolingo, ELSA Speak, voice-based chatbots (e.g., EchoDot, SmallTalk2Me), and VoiceTube have significantly boosted student engagement and motivation for consistent speaking practice. These tools are particularly popular among students due to their interactive nature, which often includes elements like rewards, levels, and challenges. By incorporating such features, these AI tools make learning more engaging and enjoyable. For example, Duolingo's integration of gamification elements, such as rewards and challenges, has been shown to enhance user engagement and motivation (Qiao & Zhao, 2023). In addition to this, Duolingo utilizes ASR and Emotional Recognition (ER) technologies, which contribute to a more engaging, game-like learning experience. These features help learners stay committed and motivated by providing immediate feedback, which is essential for language practice. Similarly, conversational AI tools, such as Enskil, have created a low-anxiety environment for learners, enabling them to practice their second language without the fear of judgment. This safe, supportive space encourages learners to engage in communication, boosting their confidence and willingness to speak (Ericsson & Johansson, 2023; Ericsson, Sofkova Hashemi & Lundin 2023). Reduced speaking anxiety Integrating AI technology into speaking classrooms has been shown to create a less stressful environment, significantly reducing learners' speaking anxiety. AI tools, such as ChatGPT and various AI chatbots, provide a non-judgmental space for learners to practice their speaking skills, which alleviates the fear of making mistakes in front of others. Sayde et al. (2024) found that ChatGPT facilitated an unbiased and supportive environment, allowing learners to engage without feeling judged. Similarly, Azizimajd (2023) emphasized that AI chatbots helped reduce speaking anxiety by providing a low-pressure platform where learners could practice at their own pace. Moreover, the adaptive nature of AI tools allows learners to move through language learning tasks based on their personal pace and needs. This individualized approach not only reduces the pressure often felt in traditional classroom settings but also fosters confidence in using new vocabulary and sentence structures. Huang and Zou (2024) pointed out that AI tools like EAP Talk enhanced learner autonomy by offering personalized learning methods and real-time feedback, contributing to a reduction in language learning anxiety and improving overall learning experiences.

4.3.3 Personalized learning and real-time feedback

AI tools offer personalized feedback tailored to individual learners' needs, helping them identify and address specific language challenges. By adapting to each student's progress, these tools provide targeted practice in areas such as grammar, vocabulary, and pronunciation. This approach enhances learning efficiency by allowing students to focus on their most pressing difficulties. According to Qiao and Zhao (2023), personalized and real-time feedback significantly improves learners' speaking proficiency by instantly identifying and correcting errors, promoting faster language acquisition. Furthermore, the adaptability of AI systems is crucial in adjusting the difficulty of tasks based on a learner's speaking proficiency. Marpaung (2024) found that Gemini AI, for instance, automatically adjusts the level of language learning tasks, providing a tailored experience that matches learners' current abilities. This personalized feedback not only helps students recognize their weaknesses but also motivates them to set clear learning objectives, fostering greater engagement and persistence in their language practice. However, although the advantages of personalized AI feedback are widely recognized, it remains to be seen whether this approach can sustainably promote the development of more complex language skills, especially in more authentic and open communicative contexts.

4.3.4 Flexibility and accessibility

The flexibility and accessibility of AI tools allow learners to practice speaking outside the classroom, offering greater autonomy in their learning process. By providing interactive exercises outside

the classroom, platforms such as Duolingo, ELSA Speak, and EAP Talk allow students to engage in independent practice and manage their own learning progress, thereby making the experience more enjoyable and personalized (Azzahra et al., 2024; Hsu et al., 2023; Zou et al., 2023a). These technologies also help overcome traditional classroom limitations, such as time constraints and the lack of individualized feedback, by offering personalized practice opportunities and timely, targeted feedback, especially in areas like pronunciation and grammar, which can significantly enhance learning efficiency and boost learners' confidence (Wang et al., 2024a). In addition to supporting learners, AI tools benefit teachers by automating routine tasks such as grading and providing immediate, detailed feedback, allowing educators to devote more attention to instructional design and classroom interaction (Yang et al., 2024). In practice, this flexibility and convenience are limited by factors such as access to devices, internet connectivity, and digital literacy. Some students still face barriers to using these tools. Therefore, as AI is increasingly applied in education, it is important to pay attention to equitable distribution of resources and support.

4.4 What are the potential limitations and challenges involved in integrating AI technologies in ESL/EFL speaking contexts?

The incorporation of AI technologies into ESL and EFL speaking environments poses various challenges and constraints, as identified through both data analysis and the review of relevant literature (Table 2).

Table 2
Limitations and Challenges for Speaking

Themes	Number of Studies Finding that Theme	Percentage (%)
Technological and accessibility	10	18%
Speech recognition and accuracy	7	13%
Lack of contextual and cultural sensitivity	6	11%
Lack social interaction	8	15%
Lack emotional engagement	7	13%

4.4.1 Inequality and accessibility issues

One of the main barriers to the adoption of AI in language speaking is the unequal access to necessary technological infrastructure, such as devices, high-speed internet, and advanced AI tools. Regarding the issue of inequality and accessibility, most of the articles collected focus on the technical limitations of AI tools themselves, such as errors in speech recognition and lack of contextual sensitivity. This issue is particularly prevalent in developing countries or under-resourced educational settings, where the digital divide prevents many students from benefiting from AI-driven learning platforms and thus exacerbates existing educational inequalities. In addition to these systemic challenges, practical constraints, such as limited internet connectivity and insufficient infrastructure in rural or economically disadvantaged regions can further restrict students' opportunities to engage in AI-based language learning. Even when technological access is available, other factors may still hinder the effective use of AI tools. For example, Saraswati et al. (2023) noted that the effective use of AI tools in language learning can be constrained by factors such as internet speed and subscription costs, which may limit access for some learners. Additionally, Ericsson and Johansson (2023) pointed out that AI learning systems sometimes experience technical issues, such as system crashes, which degrade the user experience. Furthermore, interaction with conversational agents (CAs) can be restricted by predefined pathways, resulting in inflexible and often unnatural responses that hinder effective communication.

4.4.2 Speech recognition and accuracy

AI-based language learning tools often struggle with non-native accents, dialects, and speech patterns that do not conform to standard English pronunciation, leading to inaccurate feedback, particularly for beginners or learners less familiar with these accents. This challenge undermine the effectiveness of AI as a reliable tool for language development. Many AI systems also face difficulties in accurately recognizing accents of non-native English speakers, which affects the validity of these tools in real-world language learning contexts. Sun et al. (2017) demonstrated that the use of mobile social networking applications can significantly improve EFL learners' speaking fluency, though gains in pronunciation and accuracy were more limited. They also cite prior research indicating that mobile speaking applications with ASR can support pronunciation improvement. However, speech recognition accuracy remains a challenge for AI-based intelligent personal assistants when processing non-native accented speech, often leading to communication breakdowns and a higher word error rate (Moussalli & Cardoso, 2020). Zou et al. (2023a) also highlighted that many users found the EAP Talk speech recognition system failed to accurately capture their pronunciation, leading to incorrect feedback. This issue not only affects the reliability of the system's scores but also diminishes learners' confidence in the feedback they receive. As one participant (S5) noted, "Although my pronunciation was correct, the system still flagged it as incorrect, making the feedback very unreliable." The limitations of speech recognition technology not only restrict the role of AI tools in helping learners improve their pronunciation but may also have negative psychological effects. Repeatedly receiving incorrect feedback can lead to frustration among learners and even reduce their willingness to continue using AI tools.

4.4.3 Lack of contextual and cultural sensitivity

Though AI technology is widely used in language education, its ability to capture cultural and contextual differences remains limited, which can affect the effectiveness of ESL/EFL learning. These limitations often result in unnatural or unrealistic language exercises, ultimately reducing the quality of learning outcomes. For example, Saraswati et al. (2023) found that AI tools, such as Replika, struggle to fully understand the complexity of cultural contexts. This leads to frequent misunderstandings of cultural nuances, idioms, and culturally specific expressions, which can frustrate learners, particularly when repeated miscommunications cause them to lose patience with the learning process. Similarly, Fathi et al. (2024) highlighted that AI tools struggle to recognize and respond to learners' cultural backgrounds, especially when dealing with expressions of politeness or culturally specific language use. Moreover, AI systems primarily rely on the last user input to judge intent, which limits their ability to grasp deeper meanings and contextual subtleties. At present, most research focused on the technical optimization of AI tools, while less attention is paid to how technology can be integrated with teachers and teaching materials to address shortcomings in cultural sensitivity. Given the complexity of real-world language communication, technical improvements alone are not sufficient to fully resolve these issues. Combining the strengths of AI with teachers' cultural guidance and genuine peer interaction may be the key to enhancing both the quality of language learning and its relevance to real-life situations.

4.4.4 Lack of social interaction and emotional engagement

Although AI tools can provide personalized learning, they still face significant challenges in terms of interaction and emotional engagement. Ericsson, Lundin & Hashemi (2024) found that while AI tools are convenient for language learning, they lack the ability to fully understand and respond to human emotions. This limitation prevents AI from establishing emotional connections like those formed through human interaction, potentially reducing learner motivation and engagement. Furthermore, Ullah et al. (2023) noted that, while AI chatbots like NUMLINA can effectively support English speaking practice and promote autonomous learning, they cannot fully replicate the depth of interaction or the emotional

support provided by human teachers. Although students appreciated the convenience and comfort of using chatbots, the study acknowledged that chatbots' responses are limited compared to real human communication, particularly in providing nuanced feedback and encouragement. Fathi et al. (2024) emphasized that many learners still prefer interacting with human instructors, because real teachers can offer emotional encouragement and personal support, which are important for sustaining motivation. Without this human connection, AI-based learning may lead to reduced interest and motivation over time.

4.4.5 Overreliance on AI

With the widespread integration of AI in language learning, scholars have begun to voice concerns about students' excessive reliance on AI tools. However, there is still limited attention in existing literature regarding the potential risks of such overreliance. Notably, none of the reviewed empirical studies directly examined overreliance on AI as a research focus. This indicates a significant gap in the literature and underscores the need for further research on the long-term impact of AI dependence in language learning. According to Mohebbi (2024), over-reliance on AI tools may reduce learners' engagement in critical thinking and independent problem-solving, as students risk bypassing essential cognitive processes such as analysis and reflection. Zhai et al. (2024) noted that frequent use of AI chat systems can lead to reduced critical thinking and analytical skills, as relying too much on AI may weaken independent judgment and creativity. They also pointed out that ethical concerns such as hallucinations, bias, and plagiarism are also closely associated with the risks of over-reliance. Although concerns about AI ethics are being discussed, there is still a lack of systematic research on how overreliance affects students' cognitive and academic development. While AI undoubtedly brings convenience to language learning, overreliance can undermine students' cognitive abilities, autonomy, and interpersonal skills, and may even impact academic ethics and mental health. Spatola (2024) found that when users repeatedly accept AI's direct answers, they are less likely to question recommendations or notice mistakes, especially if AI accuracy drops. This "automation complacency" means students may become passive, relying on AI instead of developing their own problem-solving skills. It is essential to remain vigilant against the misuse of this technology, strengthen the cultivation of critical thinking and independent learning, and clearly define the role of AI in education.

4.4.6 Ethical and privacy concerns

The integration of AI technologies into spoken language learning also introduces a range of ethical and privacy risks. To begin with, AI systems often require the collection and analysis of substantial amounts of learners' speech data, which involves sensitive personal information. If proper security measures are not in place during data storage and transmission, there is a risk of data breaches or unauthorized access. Moreover, some platforms do not adequately inform users about data usage or obtain clear consent, leaving learners unaware of how their information is being handled or protected. In addition, Peer Mohamed (2024) pointed out that AI algorithms may contain biases, resulting in unfair evaluations of learners based on accent, gender, or background. Vaccino-Salvadore (2023) further noted that biases in training data can be reflected in AI-generated content, with ChatGPT sometimes reproducing stereotypes and prejudices related to gender, race, or religion. For example, studies have shown that 93% of GPT-3's training data is in English, highlighting a dominance of American cultural perspectives that may undermine linguistic and cultural diversity. Such issues not only jeopardize user privacy but also impact the trustworthiness and acceptance of AI tools in educational contexts. Among these reviewed articles, many of them provide only brief descriptions of how data is handled, and privacy is safeguarded, for example, Ullah et al. (2023) described the experimental design and data analysis in detail, but didn't mention informed consent, data anonymization, ethical approval, or any privacy protection measures.

Similarly, Duong & Suppasetseree (2024) stated that "the participants were informed of research purposes, implementation, and confidentiality," and that "consent forms were administered to all participants," but didn't provide further details on how data anonymity was ensured, how the data were stored or protected, or whether any ethical approval was obtained. Such brief and general descriptions of data and privacy handling are common in the literature and may undermine transparency and replicability, which can undermine the transparency and replicability of the research. It is important to consider how ethical and privacy issues may influence research design, participant selection, data interpretation, and the scope of conclusions.

5 Discussion

5.1 Empowering speaking language learning: a theoretical interpretation of the current applications in various contexts

RQ1 systematically reviewed the application of AI technologies in ESL and EFL speaking learning from 2017 to 2024, identifying a wide range of technologies, including intelligent chatbots, ASR, speech assessment systems, and generative large language models, which have been widely integrated into various educational settings. These technologies not only expand the time and space available for speaking practice, but also fundamentally transform the learning environment. This application of AI in speaking instruction has become increasingly diversified and integrated, for example, chatbots lower the barrier to practice, IPA expand usage contexts, ASR technology enhances the accuracy of assessment and feedback, while large language models improve the realism and complexity of interactions. Specialized applications provide targeted training and automatic error correction. At the same time, this transformation marks a shift from traditional reliance on tools to the participation of collaborative tools and platforms, which advances the development of knowledge construction theories (Scardamalia & Bereiter, 2010; Vygotsky, 1978). However, the deep integration of AI has also brought about several significant theoretical and practical challenges. On the one hand, the role of teachers is undergoing transformation, shifting from knowledge transmitters to facilitators and guides of learning, which requires higher levels of professional expertise and technological adaptability. On the other hand, learners' autonomy and critical thinking skills may be weakened under the convenience of highly intelligent support, as over-reliance on technological tools could undermine their initiative for exploration. In addition, the digital divide and unequal resource distribution not only limit access to AI-based educational resources for certain groups, but also further exacerbate issues of educational equity (Crompton et al., 2024). It is also worth noting that while AI-driven learning environments improve efficiency and interactivity, whether they can truly replace the social and emotional support gained from interpersonal interactions remains to be fully examined.

Therefore, the current application of AI in ESL and EFL speaking learning not only demonstrates the immense potential for technological innovation, but also reveals complex issues, including concerns about data privacy and security, educational equity, resource allocation, ethical challenges, and the need to support diverse learner needs. In the future, as AI technology continues to develop and integrate, its role in speaking education will become increasingly diverse and indispensable, yet it will also pose higher standards and new challenges for teacher professional development, curriculum design, learning objectives, and assessment systems. There is an urgent need to seek a balance between technological empowerment and educational values, to promote an AI-enabled language learning paradigm that is both innovative and inclusive, ultimately fostering a more equitable educational ecosystem.

5.2 Learning mechanisms and methodological innovations driven by intelligent technology

Based on the analysis in RQ2, research on the application of AI technology in spoken language teaching mainly focuses on higher education and formal educational settings, with most studies targeting

intermediate and advanced learners. Mixed research designs have become mainstream. However, there are significant limitations in the samples and contexts: most studies have small sample sizes, short durations, and lack cultural and regional diversity. Furthermore, there is a lack of long-term follow-up or in-depth analysis of beginners, learners at different levels, and those from multicultural backgrounds. These narrow focuses result in three main issues:

First, small sample sizes and short durations limit the statistical significance and external validity of the findings (Etz & Arroyo, 2015). These studies fail to effectively capture long-term learning processes and cannot account for learners from diverse backgrounds with varying needs, which reduces the generalizability of the conclusions and makes them less applicable to other contexts. Second, existing research places excessive emphasis on intermediate learners, which limits the generalizability and applicability of the results (Ercikan & Roth, 2014). As a result, conclusions are difficult to apply to beginners, advanced learners, or heterogeneous groups. Moreover, learners in transitional or marginal stages (such as low-level or mixed-level groups) are underrepresented, even though these groups often require more targeted support. The focus on learners with stronger foundational abilities exacerbates the unequal distribution of educational resources (Lu & Matheny, 2025), placing beginners and learners from diverse backgrounds at a disadvantage when accessing technological innovations. Finally, the lack of cultural and regional diversity restricts the ability to identify global or cross-cultural issues among learners. This bias tends to overlook the unique needs and challenges of learners from different cultural or socioeconomic backgrounds, further deepening the unequal distribution of educational resources. Additionally, the absence of long-term tracking makes it difficult to understand learners' progress at various stages. The lack of in-depth analysis hinders the development of targeted educational technology strategies that could better assist low-level or multicultural learners.

This situation suggests that research on educational technology requires methodological innovation and context diversification. Future studies should not only expand to cross-cultural, cross-grade, and longitudinal research, but also strengthen collaboration among diverse stakeholders, including teachers, students, policymakers, educational technologists, and researchers, and enhance contextual accuracy by carefully considering various educational settings, learner backgrounds, and instructional contexts. This approach helps prevent the benefits of technology from being overstated or evaluated too narrowly, ensuring that research findings are relevant and applicable to real-world educational environments. Besides that, expanding research perspectives is essential for achieving the synergistic innovation of technology and learning mechanisms, providing a more solid foundation for the empirical and theoretical development of technology in speaking education.

5.3 The mechanisms of AI in enhancing speaking skills and theoretical elevation

RQ3 showed that AI technology improves ESL/EFL learners' speaking skills in several ways. Using big data and adaptive algorithms, it provides personalized, real-time feedback and creates a low-anxiety environment that boosts motivation and confidence in speaking (Du & Daniel, 2024). AI also extends the time and space available for speaking practice, enhancing learners' autonomy and persistence. This goes beyond just improving technical efficiency. AI also allows for the reinterpretation and expansion of learning theories in the digital age. However, the sustainability of these improvements, especially in developing complex language use and cross-cultural communication skills, still require guidance from human teachers and collaboration among various participants. The improvement in speaking should not just be seen as "skill growth" under technology but as a fundamental restructuring of language ability. With AI's integration, speaking a language evolves from a static skill into a dynamic, social, and multimodal process of meaning-making (Ding & Yusof, 2025; Zou et al., 2023b). In this process, learners are no longer just "receivers" or "imitators" but active negotiators of meaning, creators of conversational contexts, and builders of multiple identities. The combination of technology and language learning expands the boundaries of language skills, linking them with cultural sensitivity, multimodal expression,

and critical thinking. Language, in this context, becomes not only a tool for communication but also a platform for learners' self-awareness, emotional regulation, and social participation.

5.4 Complexities and concerns in AI-Assisted speaking education

RQ4 reviewed challenges using AI tools into language learning. While AI can efficiently simulate conversation, it struggles to understand context, tone, and cultural expressions. Its feedback, based on algorithms and big data, tends to evaluate student responses in a standardized way, overlooking individual creativity and uniqueness. This can lead students to focus on "matching AI feedback" rather than fostering critical thinking in language use (Nguyen, 2024; Zhai et al., 2024). Additionally, AI cannot accurately interpret learners' emotions, non-verbal cues, or complex communication strategies, which make interactions feel mechanical and less adaptable. If language practice involves interacting only with a machine for long periods, learners may struggle to feel a sense of belonging or understanding. Most existing research tends to focus on improving the functional capabilities of AI tools, while discussions on how to effectively foster emotional connections between teachers and students or among peers with the support of AI remain limited, it is important to continue emphasizing the irreplaceable value of emotional support and genuine human interaction, to achieve a more comprehensive and sustainable learning experience.

Privacy concerns also hinder AI's widespread use in education. Some learners avoid participating in AI-assisted speaking practice or studies due to fears of data breaches, affecting the diversity and representativeness of research samples. As Li et al. 2025 noted, students are worried that collecting and using their personal data could lead to privacy leaks and a loss of trust. Ethical reviews and data protection may further limit researchers' ability to observe the real learning process, impacting the completeness of data. Therefore, AI should not be seen as the ultimate solution for speaking education but as a supportive tool that complements human interaction and cultural experience. In the future, AI should not operate in isolation but be integrated with teachers' instructional guidance, peer-to-peer interaction, and authentic social contexts. For example, AI-powered speaking tools can support teachers by providing personalized feedback on students' oral performance, facilitate group discussions and collaborative projects among peers, and simulate real-world conversational scenarios that mirror everyday communication. By combining the strengths of AI with human support and real-life practice, students can receive more comprehensive language input, develop practical speaking skills, and gain greater confidence in using English outside the classroom.

In addition, after reviewing these articles, we found that many existing studies (e.g., Jeon & Lee, 2024; Khailifa & Ginting, 2024; Sayed et al., 2024; Ye et al., 2022) tend to be overly optimistic in their evaluation of AI tools, often overlooking its potential side effects. Literature review revealed that some scholars rarely conduct in-depth analyses of the problems and challenges that AI tools may encounter in practical applications. Some negative impacts or potential risks are often glossed over, lacking empirical research and rigorous data support. For instance, Jeon and Lee (2024) highlighted the positive effects of chatbot-assisted flipped learning on EFL learners' motivation, interaction, and performance. However, the study discussed methodological limitations only and scarcely addressed the potential negative impacts or side effects of AI technologies in education. This research tendency can easily lead to blind advocacy for AI technology and unrealistic expectations. Therefore, future research should not only recognize the advantages of AI but also strengthen systematic analyses of its risks and challenges to achieve more comprehensive and balanced development.

5.5 Future research: innovation, ethical governance, and theoretical expansion

The application of AI in ESL/EFL speaking learning needs to be strengthened in the areas of equity, emotional intelligence, diversity, and ethical governance. First, educational equity should be a core

focus in AI design, especially in low-resource environments. It is important to develop low-cost, low-bandwidth tools to ensure all learners have equal access to educational resources and reduce the digital divide. For instance, offline modes or lightweight applications could be developed in regions with inadequate internet infrastructure. Second, AI should focus on improving social-emotional intelligence and cultural sensitivity (Sethi & Jain, 2024). AI tools need to be adapted to specific cultural contexts to enhance interaction quality and reduce misunderstandings caused by cultural differences. Third, the collaboration between AI, teachers, and peers is key to improving learning outcomes (Semerikov et al., 2021; Yang, 2024). AI should complement teacher guidance and peer interaction, offering emotional support and cultural recognition, rather than relying solely on AI assessments. Finally, ethical governance is essential, and future AI applications must have strict standards for privacy protection, algorithm transparency, and fairness to ensure data is used legally and without bias (Peer Mohamed, 2024).

As highlighted by Dwivedi et al. (2023) and Holmes et al. (2022), achieving a balance between technological advancement and humanistic values, such as equity, empathy, and ethical awareness is essential for AI to truly foster diversity, inclusion, and sustainable development in education. Future research should focus more on incorporating AI applications into sociocultural theory, critical educational technology theory, and other frameworks. Understanding AI's impact on learners' cognition, identity, and social relationships will help clarify its role in education.

6 Limitations of This Review

Although this review aims to provide a comprehensive overview of research on AI-supported speaking instruction worldwide, several limitations should be acknowledged. First, there is a regional imbalance in the literature: among the 55 included studies, 19 are from China, which may affect the generalizability of the findings. This is because educational systems, classroom practices, technology infrastructure, and learners' attitudes toward AI can vary greatly between countries. As a result, findings based largely on Chinese contexts may not be directly applicable to other regions with different sociocultural and educational backgrounds. Additionally, most of the reviewed studies focus on EFL contexts, with little attention given to ESL environments. This limits our understanding of how AI tools perform across different language learning settings. Furthermore, this study included only 55 articles as the objects of analysis, resulting in a limited sample size. This may affect the representativeness of the findings, and some important perspectives and recent research developments might have been overlooked. Lastly, the selection process also emphasized high-impact journals and methodological rigor, potentially overlooking innovative practices and emerging trends. Future research should address these gaps by including studies from a wider range of regions and language contexts, and by incorporating both qualitative and longitudinal research designs to gain a deeper understanding of the mechanisms and impact of AI-supported speaking instruction. In addition, expanding the literature sample will help ensure that conclusions are more comprehensive and reliable.

7 Conclusion

This study systematically reviewed the AI-based applications, benefits, and barriers of AI in improving ESL/EFL speaking skills, providing valuable insights into its transformative potential. This review not only maps the current applications of AI in ESL/EFL speaking instruction but also offers new perspectives on how technology shapes learners' expression, confidence, and learning habits. By systematically analyzing different AI tools, learning contexts, and research approaches, the study highlights both the opportunities and complexities that come with integrating technology and language teaching. It also demonstrates the potential and limitations of human–AI collaboration in promoting motivation, personalized development, and intercultural communication.

Different from previous studies, our review makes a unique contribution by systematically identifying and analyzing several practical issues that have received little attention in previous studies on AI-supported ESL/EFL speaking instruction. While confirming that AI can help increase students' participation and speaking confidence, this review also finds that most AI tools use the same practice materials and scoring standards for all learners (Chen & Sun, 2025; Zou et al., 2023a). As a result, many students feel the feedback is too generic and does not meet their individual needs. The review further notes that teachers' roles are changing as they need to adapt to new technologies, and some teachers feel uncertain about how best to use these tools (Alwaqdani, 2025; Nazim & Alzubi 2025). Concerns about the accuracy of automatic feedback and the protection of student data are also becoming more prominent. In addition, while AI can help students communicate in different cultural settings and encourage independent learning, it may also create some obstacles in these areas. By highlighting these real classroom challenges, this review not only adds to current knowledge, but also offers practical suggestions for designing better AI-based speaking activities. These findings show the importance of ongoing research to test and improve the use of AI in real teaching settings.

More importantly, this review underlines the need to continuously reflect on the core aims of education and the role of learners in a rapidly changing technological environment. It is hoped that the insights presented here can provide a theoretical basis for future research and practice in AI-supported speaking instruction, as well as encourage further interdisciplinary exploration in this field. While technology offers new possibilities, meaningful language learning will always depend on ongoing inquiry and dialogue.

Appendix

A1 Contextual features of all the included studies

	Author (Year)	AI technology	Study Design	Sample Size	Study Setting	Learner type	Context/ Country
JA01	Ericsson et al. (2023)	AI Chatbots-Enskill	Mixed	25	Formal learning	EFL learners, A1–A2	Sweden
JA02	Hsu et al. (2023A)	Task-oriented chatbot system (TPBOT – TOEIC Practice Chatbot)	Quantitative	Not directly stated	Informal learning	EFL learners, A1–A2	Taiwan (China)
JA03	Junaidi et al. (2020)	IPA-Lyra Virtual Assistant (LVA)	Quantitative	65	Formal learning	EFL learners, A1–A2	Indonesia
JA04	Sun et al. (2017)	Mobile Social Networking Site (SNS) "Papa"	Mixed	72	Formal learning	EFL learners, pre-A1–A1	China
JA05	Kim et al. (2021)	AI chatbots (Replika, Andy, Google Assistant)	Mixed	49	Formal learning	EFL learners, A2–B1	South Korea
JA06	Azzahra et al. (2024)	Duolingo (ASR) and ELSA Speak	Qualitative	6	Formal learning	EFL learners, B1–B2	Indonesia
JA07	Gebregziabher et al. (2024)	Intelligent voice learning system based on AI wireless network	Mixed	Not specified	Formal learning	EFL learners, A1–A2	China

	Author (Year)	AI technology	Study Design	Sample Size	Study Setting	Learner type	Context/ Country
JA08	Khalizah & Damanik (2024)	ELSA Speak	Mixed	23	Formal learning	EFL learners A1–A2	Indonesia
JA09	Hsu et al. (2023B)	IPA	Quantitative	Two classes	Formal learning	EFL learners, B1–B2	Taiwan (China)
JA10	Fathi et al. (2024)	Andy English Chatbot	Mixed	65	Formal learning	EFL learners, B1–B2	Iran
JA11	Annamalai et al. (2023)	Chatbots	Mixed	360	Formal learning	EFL learners, B1–B2	Malaysia
JA12	Ericsson & Johansson. (2023)	virtual humans in Enskill (spoken dialogue system)	Qualitative	22	Formal learning	EFL learners, A1–A2	Sweden
JA13	Ahyarudin & Jamilah (2024)	AI Chatbots, AI-Text-To-Speech & AI Speech-To-Text	Qualitative	35	Formal learning	EFL learners, A2–B1	Indonesia
JA14	Ullah et al. (2023)	NUMLINA chatbot	Quantitative	30	Formal learning	EFL learners, B2	Pakistan
JA15	Dizon. (2020)	IPA-Alexa	Quantitative	28	Formal learning	EFL learners, A1–B2	Japan
JA16	Zou et al. (2024)	EAP Talk-ASR	Mixed	366	Formal learning	EFL learners, B1–B2	China
JA17	El Shazly (2021)	chatbots	Mixed	48	Formal learning	EFL learners, B1–B2	Egypt
JA18	Yang et al. (2024)	AI Chatbots-TalkAI	Mixed	36	Formal learning	EFL learners, B1–B2	China
JA19	Pebriani et al. (2024)	Unspecified AI tools (general AI use in EFL speaking class)	Quantitative	40	Formal learning	EFL learners, B1	Indonesia
JA20	Üstünbaş (2024)	ChatGPT	Qualitative	4	Formal learning	EFL learners, B1	Türkiye
JA21	Saraswati et al. (2023)	AI Chatbots-AI Replika	Qualitative	39	Formal learning	EFL learners, B1	Indonesia
JA22	Marpaung (2024)	AI Chatbots-Gemini AI	Mixed	30	Formal learning	EFL learners, A2–B2	Indonesia
JA23	Duong & Suppasetseree (2024)	Andy English Bot	Mixed	30	Formal learning	EFL learners, A2	Vietnam
JA24	Aldosari (2024)	ChatGPT	Quantitative	73	Formal learning	EFL learners, B1–B2	Saudi Arabia
JA25	Zou et al. (2023A)	AI speech evaluation programs	Mixed	40	Informal learning	EFL learners, B1–B2	China

	Author (Year)	AI technology	Study Design	Sample Size	Study Setting	Learner type	Context/ Country
JA26	Qiao & Zhao (2023)	Duolingo	Quantitative	93	Formal learning	EFL learners, B1–B2	China
JA27	Çakmak (2022)	Chatbot Replika (AI-based conversational agent)	Mixed	90	Formal learning	EFL learners, B1	Turkey
JA28	Chen et al. (2022)	DA-SR, CF-SR	Quantitative	56	Formal learning	EFL learners, A1	Taiwan (China)
JA29	Warman et al. (2023)	ELSA Speak, Duolingo, and Orai	Quantitative	85	Formal learning	EFL learners, A2–B1	Indonesia
JA30	Ulinuha & Parnawati (2024)	SmallTalk2Me	Qualitative	78	Formal learning	EFL learners, most B1–B2, some A2/C1	Indonesia
JA31	Jeon & Lee (2024)	Customized chatbots	Mixed	87	Formal learning	EFL learners, A1 -A2	South Korea
JA32	Lin & Mubarok (2021)	Replika	Quantitative	50	Formal learning	EFL learners, B1–B2	Taiwan (China)
JA33	Muniandy & Selvanathan (2024)	ChatGPT	Mixed	40	Formal learning	ESL learners, B1, C1	Malaysia
JA34	Huang & Zou (2024)	EAP Talk	Quantitative	203	Formal learning	EFL learners, B1–B2	China
JA35	Ye et al. (2022)	Microsoft Xiaoying	Quantitative	56	Formal learning	EFL learners, A2–B1	China
JA36	Bashor et al. (2022)	ASR	Mixed	573	Formal learning	EFL learners, A2-B1	Indonesia
JA37	Zou et al. (2023B)	AI speech evaluation apps	Mixed	70	Formal learning	EFL learners, B1-B2	China
JA38	Wang et al. (2024)	Conversational GenAI Chatbots	Mixed	99	Formal learning	EFL learners, B1-B2	China
JA39	Zhang et al. (2024)	AI Chatbot (iFLYTEK Spark), ITS (FiF Speaking)	Mixed	82	Formal learning	EFL, B1-B2	China
JA40	Ericsson et al. (2024)	Conversational AI (SDS, Enskill)	Mixed	22	Formal learning	EFL, A1–A2	Sweden
JA41	Tai (2024)	IPA (Google Assistant)	Mixed	89	Informai learning	EFL, B1–B2	Taiwan (China)
JA42	Nguyen et al. (2024)	AI-based platforms: SmallTalk2Me, VoiceTube	Quantitative	108	Formal learning	EFL, A2–B1	Vietnam

	Author (Year)	AI technology	Study Design	Sample Size	Study Setting	Learner type	Context/ Country
JA43	Han (2020)	Voice-based AI chatbot (Echodot)	Quantitative	44	Formal learning	EFL, A1-A2	Korea
JA44	Khailifa & Ginting (2024)	General AI platforms for speaking enhancement	Mixed	15	Formal learning	EFL, B1-B2	Indonesia
JA45	Yang et al. (2024)	IPA: Google Assistant, delivered via Google Nest Mini smart speaker	Mixed	34	Formal learning	EFL, B1–B2	Taiwan (China)
JA46	Yang et al. (2022)	Task-based voice chatbot (Ellie)	Quantitative	314	Formal learning	EFL, A1–A2	South Korea
JA47	Muthmainnah (2024)	AI-CiciBot (Conversational AI)	Mixed	30	Formal learning	EFL, A2–B1	Indonesia
JA48	Kholis (2021)	ELSA Speak	Mixed	18	Formal learning	EFL, B1–B2	Indonesia
JA49	Moussalli, S., & Cardoso, W. (2020)	Intelligent Personal Assistant (IPA) - Alexa (Amazon Echo)	Mixed	11	Formal learning	ESL, B1–C1	Canada
JA50	Dew et al. (2022)	Chatbot (Memrise)	Quantitative	36	Formal learning	EFL, A2–B1	Indonesia
JA51	Tai & Chen (2020)	(IPA) - Google Assistant	Mixed	112	Formal learning	EFL learners, A2–B1	Taiwan (China)
JA52	Zhang (2024)	(IPA) - Google Assistant	Mixed	54	Formal learning	EFL, B1	China
JA53	Ibrahim et al. (2024)	Intelligent Computer-Assisted Language Assessment (ICALA)	Quantitative	81	Formal learning	EFL, B1	Saudi Arabia
JA54	Sayed et al. (2024)	AI-Powered Tools (ChatGPT)	Mixed	28	Formal learning	EFL, B2	Ethiopia
JA55	Azizimajd (2023)	Voice-based Chatterbot (Replika)	Quantitative	60	Formal learning	EFL, B2	Iran

References

Annamalai, N., Rashid, R. A., Hashmi, U. M., Mohamed, M., Alqaryouti, M. H., & Sadeq, A. E. (2023). Using chatbots for English language learning in higher education. *Computers and Education Artificial Intelligence*, *5*, 100153. https://doi.org/10.1016/j.caeai.2023.100153

Aziz, A. A., & Kashinathan, S. (2021). ESL learners' challenges in speaking English in Malaysian classroom. *Development*, *10*(2), 983-991. https://doi.org/10.6007/ijarped/v10-i2/10355

Azizimajd, H. (2023). Investigating the impacts of voice-based student-chatbot interactions in the classroom on EFL learners' oral fluency and foreign language speaking anxiety. *Technology Assisted Language Education*, 1(1), 61–83. https://doi.org/10.22126/tale.2023.2732

Azzahra, N., Widiastuti, P. T., Sopyani, N., Luthfiyyah, R., & Dwiniasih, N. (2024). A study on English students' perception regarding the AI implementation on listening and speaking skills. *Indonesian Journal of Innovation and Applied Sciences (IJIAS)*, *4*(2), 113–120. https://doi.org/10.47540/ijias.v4i2.1466

Alwaqdani, M. (2025). Investigating teachers' perceptions of artificial intelligence tools in education: Potential and difficulties. *Education and Information Technologies, 30*(3), 2737–2755. https://doi.org/10.1007/s10639-024-12903-9

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

Chang, W. L., & Sun, J. C. Y. (2024). Evaluating AI's impact on self-regulated language learning: A systematic review. *System*, *126*, 103484. https://doi.org/10.1016/j.system.2024.103484

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13460

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00392-8

Chen, T., & Sun, S. (2025). Evaluating automated evaluation systems for spoken English proficiency: An exploratory comparative study with human raters. *PloS one*, *20*(3), e0320811. https://doi.org/10.1371/journal.pone.0320811

Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology, 24*(1), 16–26. https://doi.org/10125/44705

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education Artificial Intelligence*, *6*, 100230. https://doi.org/10.1016/j.caeai.2024.100230

Duong, T., & Suppasetseree, S. (2024). The effects of an artificial intelligence voice chatbot on improving Vietnamese undergraduate students' English speaking skills. *International Journal of Learning Teaching and Educational Research*, *23*(3), 293–321. https://doi.org/10.26803/ijlter.23.3.15

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). Opinion Paper:"So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International journal of information management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Ding, D., & Yusof, M. B. (2025). Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: a mixed methods study. *Humanities and Social Sciences Communications*, *12*(1), 1-16. https://doi.org/10.1057/s41599-025-05550-z

Ercikan, K., & Roth, W. (2014). Limits of generalizing in education research: Why Criteria for research generalization should include population heterogeneity and uses of knowledge claims. *Teachers College Record the Voice of Scholarship in Education*, *116*(4), 1–28. https://doi.org/10.1177/016146811411600405

Ericsson, E., & Johansson, S. (2023). English speaking practice with conversational AI: Lower secondary students' educational experiences over time. *Computers and Education Artificial Intelligence*, *5*, 100164. https://doi.org/10.1016/j.caeai.2023.100164

Ericsson, E., Lundin, J., & Hashemi, S. S. (2024). From deadpan machine to relating socially: Middle school students' experiences speaking English with embodied conversational agents. *Journal of Research on Technology in Education*, *56*(6), 752–768. https://doi.org/10.1080/15391523.2023.2213902

Ericsson, E., Sofkova Hashemi, S., & Lundin, J. (2023). Fun and frustrating: Students' perspectives on practising speaking English with virtual humans. *Cogent Education*, *10*(1), 2170088. https://doi.org/10.1080/2331186X.2023.2170088

Etz, K. E., & Arroyo, J. A. (2015). Small sample research: Considerations beyond Statistical power. *Prevention Science*, *16*(7), 1033–1036. https://doi.org/10.1007/s11121-015-0585-4

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, *121*, 103254. https://doi.org/10.1016/j.system.2024.103254

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, scope, Limits, and Consequences. *Minds and Machines*, *30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Hao, Z., Fang, F., & Peng, J.-E. (2024). The integration of AI technology and critical thinking in English major education in China: Opportunities, challenges, and future prospects. *Digital Applied Linguistics, 1,* 2256. https://doi.org/10.29140/dal.v1.2256

Hsu, H. L., Chen, H. H. J., & Todd, A. G. (2023). Investigating the impact of the Amazon Alexa on the development of L2 listening and speaking skills. *Interactive Learning Environments*, *31*(9), 5732-5745. https://doi.org/10.59435/gjmi.v2i2.278

Huang, F., & Zou, B. (2024). English speaking with artificial intelligence (AI): The roles of enjoyment, willingness to communicate with AI, and innovativeness. *Computers in Human Behavior*, 108355. https://doi.org/10.1016/j.chb.2024.108355

Huang, J. (2024). Enhancing EFL Speaking Feedback with ChatGPT's Voice Prompts. *International Journal of TESOL Studies*, *6*(3). https://doi.org/10.58304/ijts.20240302

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., ... & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, *32*(3), 504-526. https://doi.org/10.1007/s40593-021-00239-1

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., ... & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for information*, *34*(4), 285-291. https://doi.org/10.3233/EFI-180221

Imran, M. C., Amaliah, N., Syam, N. I., Room, F., & Sage, M. S. (2023). The feasibility of artificial intelligence (AI) in speaking skill: Lecturers' perceptions. *IJOLEH: International Journal of Education and Humanities, 2*(2), 135–144. https://doi.org/10.56314/ijoleh.v2i2

Jiang, R. (2022). How does artificial intelligence empower EFL teaching and learning nowadays? A review on artificial intelligence in the EFL context. *Frontiers in Psychology*, *13*, 1049401. https://doi.org/10.3389/fpsyg.2022.1049401

Jeon, J., & Lee, S. (2024). The impact of a chatbot-assisted flipped approach on EFL learner interaction. *Educational Technology & Society*, *27*(4), 218-234. https://doi.org/10.30191/ETS.202410_27(4).RP12

Kang, H. (2022). Effects of Artificial Intelligence (AI) and Native Speaker Interlocutors on ESL Learners' Speaking Ability and Affective Aspects. *Multimedia-Assisted Language Learning*, *25*(2). https://doi.org/10.15702/mall.2022.25.2.09

Kartal, G., & Yeşilyurt, Y. E. (2024). A bibliometric analysis of artificial intelligence in L2 teaching and applied linguistics between 1995 and 2022. *ReCALL*, *36*(3), 359–375. https://doi.org/10.1017/s0958344024000077

Khailifa, N. F., & Ginting, P. (2024). Exploring EFL Students Cognition and Practices in Applying English Speaking Through the Use of AI (Artificial Intelligence). *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, *12*(1), 417-440.

Kholis, A. (2021). Elsa speak app: automatic speech recognition (ASR) for supplementing English pronunciation skills. *Pedagogy: Journal of English Language Teaching*, *9*(1), 01-14. https://doi.org/10.32332/joelt.v9i1.2723

Kim, H. S., Cha, Y., & Kim, N. Y. (2021). Effects of AI chatbots on EFL students' communication skills. 영어학, *21*, 712-734.

Koç, F. Ş., & Savaş, P. (2024). The use of artificially intelligent chatbots in English language learning: A systematic meta-synthesis study of articles published between 2010 and 2024. *ReCALL*, 1–18. https://doi.org/10.1017/s0958344024000168

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology* (3rd ed.). Sage Publications.

Li, Y., Zhou, X., & Chiu, T. K. (2025). Systematic review on artificial intelligence chatbots and ChatGPT for language learning and research from self-determination theory (SDT): What are the roles of teachers? *Interactive Learning Environments, 33*(3), 1850–1864. https://doi.org/10.1080/10494820.2024.2400090

Liu, G. L., Lee, J. S., & Zhao, X. (2025). Critical Digital Literacies, Agentic Practices, and AI-mediated Informal Digital Learning of English. *System*, 103797. https://doi.org/10.1016/j.system.2025.103797

Liu, Y., Zhang, H., Jiang, M., Chen, J., & Wang, M. (2024). A systematic review of research on emotional artificial intelligence in English language education. *System*, *126*, 103478. https://doi.org/10.1016/j.system.2024.103478

Lu, A., & Matheny, K. T. (2025). Equal inputs, unequal outputs: How capacity limits policy implementation. *Educational Evaluation and Policy Analysis*. https://doi.org/10.3102/01623737241310832

Liang, J. C., Hwang, G. J., Chen, M. R. A., & Darmawansah, D. (2023). Roles and research foci of artificial intelligence in language education: An integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments*, *31*(7), 4270-4296. https://doi.org/10.1080/10494820.2021.1958348

Li, Y., Castulo, N. J., & Xu, X. (2025). Embracing or rejecting AI? A mixed-method study on undergraduate students' perceptions of artificial intelligence at a private university in China. *Frontiers in Education, 10*, Article 1505856. https://doi.org/10.3389/feduc.2025.1505856

Madhavi, E., Sivapurapu, L., Koppula, V., Rani, P. E., & Sreehari, V. (2023). Developing learners' English-speaking skills using ICT and AI tools. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *32*(2), 142-153. https://doi.org/10.37934/araset.32.2.142153

Marpaung, S. (2024). Boosting high school students' speaking proficiency with Gemini AI. *English Journal Antartika*, *2*(1), 8-15. https://doi.org/10.70052/eja.v2i1.543

Moussalli, S., & Cardoso, W. (2020). Intelligent personal assistants: can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning*, *33*(8), 865-890. https://doi.org/10.1080/09588221.2019.1595664

Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC medical research methodology*, *18*, 1-7. https://doi.org/10.1186/s12874-018-0611-x

Mohebbi, A. (2024). Enabling learner independence and self-regulation in language education using AI tools: a systematic review. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186x.2024.2433814

Nguyen, X. H., Le, T. T., Nguyen, T. T. H., Do Dang Khoi, H. Q. N., & Hoang, Q. N. (2024). Students' attitudes toward utilizing AI-based technologies to improve their speaking proficiency. *Ho Chi Minh City University of Education Journal of Science*, *21*(5), 957–970. https://doi.org/10.54607/hcmue.js.21.5.4270

Nazim, M., & Alzubi, A. A. F. (2025). Empowering EFL teachers' perceptions of generative AI-mediated self-professionalism. *PLoS One*, *20*(6), e0326735. https://doi.org/10.1371/journal.pone.0326735

Peer Mohamed, M. S. (2024). Exploring ethical dimensions of AI-enhanced language education: A literature perspective. *Technology in Language Teaching & Learning, 6(3), 89–99*. https://doi.org/10.29140/tltl.v6n3.1813

Pham, V. T. T., & Pham, A. T. (2025). English major students' satisfaction with ELSA Speak in English pronunciation courses. *PLoS ONE*, *20*(1), e0317378. https://doi.org/10.1371/journal.pone.0317378

Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology*, *14*, 1255594. https://doi.org/10.3389/fpsyg.2023.1255594

Rajendran, T., & Yunus, M. M. (2021). A systematic literature review on the use of mobile-assisted language Learning (MALL) for enhancing speaking skills among ESL and EFL learners. *International Journal of Academic Research in Progressive Education and Development*, *10*(1), 586-609. http://dx.doi.org/10.6007/IJARPED/v10-i1/8939

Saraswati, G. P. D., Farida, A. N., & Yuliati, Y. (2023, November). Implementing AI Replika in higher education speaking classes: Benefits and challenges. In *ELT Forum: Journal of English Language Teaching* (Vol. 12, No. 3, pp. 207-215). https://doi.org/10.15294/elt.v12i3.76525

Scardamalia, M., & Bereiter, C. (2010). A brief history of knowledge building. *Canadian Journal of Learning and Technology / La Revue Canadienne de l'Apprentissage et de La Technologie*, *36*(1). https://doi.org/10.21432/t2859m

Scells, H., Zuccon, G., Koopman, B., & Clark, J. (2020). Automatic Boolean query formulation for systematic review literature search. *In Proceedings of the Web Conference 2020*. https://doi.org/10.1145/3366423.3380185

Semerikov, S. O., Striuk, A. M., & Shalatska, H. M. (2021). AI-assisted language education: critical review. *Educational Dimension*, *4*, 1–7. https://doi.org/10.31812/ed.623

Sethi, S. S., & Jain, K. (2024). AI technologies for social emotional learning: recent research and future directions. *Journal of Research in Innovative Teaching & Learning*, *17*(2), 213–225. https://doi.org/10.1108/jrit-03-2024-0073

Sun, Z., Lin, C., You, J., Shen, H. J., Qi, S., & Luo, L. (2017). Improving the English-speaking skills of young learners through mobile social networking. *Computer Assisted Language Learning*, *30*(3–4), 304–324. https://doi.org/10.1080/09588221.2017.1308384

Sayed, B. T., Bani Younes, Z. B., Alkhayyat, A., Adhamova, I., & Teferi, H. (2024). To be with artificial intelligence in oral test or not to be: a probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. *Language Testing in Asia*, *14*(1), 49. https://doi.org/10.1186/s40468-024-00321-0

Sherwood, L., & Mac Donald, K. Leveraging a diagnostic exam for cultivating AI literacy: postgraduate student reflections on ChatGPT. (2024). *International Journal of TESOL Studies*. https://doi.org/10.58304/ijts.20240307

Spatola, N. (2024). The efficiency-accountability tradeoff in AI integration: Effects on human performance and over-reliance. *Computers in Human Behavior: Artificial Humans*, *2*(2), 100099. https://doi.org/10.1016/j.chbah.2024.100099

Toncelli, R., & Kostka, I. (2024). A love-hate relationship: Exploring faculty attitudes towards GenAI and its integration into teaching. *International Journal of TESOL Studies*, *6*(3), 77-94. https://doi.org/10.58304/ijts.20240306

Teng, M. F. (2023). Scientific writing, reviewing, and editing for open-access TESOL journals: The role of ChatGPT. *International Journal of TESOL Studies*, *5*(1), 87-91. https://doi.org/10.58304/ijts.20230107

Teng, M. F. (2024). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies*, *6*(3), 36-57. https://doi.org/10.58304/ijts.20240304

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, *8*(1). https://doi.org/10.1186/1471-2288-8-45

Ullah, Z., Ullah, S., Khalid, M. N., & Abbas, M. M. (2023). Measuring efficacy of speaking English chatbot NUMLINA: A user study. *Al-Qanṭara*, *9*(4), 369-387.

Ulla, M. B., & Teng, M. F. (2024). Generative artificial intelligence (AI) applications in TESOL: Opportunities, issues, and perspectives. *International Journal of TESOL Studies*, *6*(3), 1-3. https://doi.org/10.58304/ijts.20240301

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. *Harvard University Press*.

Vaccino-Salvadore, S. (2023). Exploring the ethical dimensions of using ChatGPT in language learning and beyond. Languages, 8(3), 191. https://doi.org/10.3390/languages8030191

Wang, C., Zou, B., Du, Y., & Wang, Z. (2024a). The Impact of different conversational generative AI chatbots on EFL Learners: an analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System*, 103533–103533. https://doi.org/10.1016/j.system.2024.103533

Wang, F., Cheung, A. C., & Chai, C. S. (2024b). Language learning development in human-AI Interaction: A thematic review of the research landscape. *System*, 103424. https://doi.org/10.1016/j.system.2024.103424

Yang, F., Li, K., & Li, R. (2024). AI in language education: Enhancing learners' speaking awareness through AI-Supported training. *International Journal of Information and Education Technology*, *14*(6), 828-833.

Yamaoka, K. (2024). ChatGPT's motivational effects on Japanese university EFL learners: A qualitative analysis. *International Journal of TESOL Studies*, *6*(3). https://doi.org/10.58304/ijts.20240303

Ye, Y., Deng, J., Liang, Q., & Liu, X. (2022). Using a smartphone-based chatbot in EFL learners' oral tasks. *International Journal of Mobile and Blended Learning*, *14*(1), 1–17. https://doi.org/10.4018/ijmbl.299405

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International journal of educational technology in higher education*, *16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, *11*(28), 1–37. https://doi.org/10.1186/s40561-024-00316-7

Zhang, D., Wu, J. G., & Fu, Z. (2024). From shy to fly: Facilitating EFL learners' willingness to communicate with an AI chatbot and an intelligent tutoring system. *System*, *127*, 103501. https://doi.org/10.1016/j.system.2024.103501

Zhu, M., & Wang, C. (2025). A systematic review of research on AI in language education: Current status and future implications. *Language Learning & Technology*, *29*(1), 1-29. https://hdl.handle.net/10125/73606

Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023a). An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *Sage Open*, *13*(3), 21582440231193818. https://doi.org/10.1177/21582440231193818

Zou, B., Liviero, S., Hao, M., & Wei, C. (2020). Artificial intelligence technology for EAP speaking skills: Student perceptions of opportunities and challenges. *Technology and the psychology of second language learners and users*, 433-463. https://doi.org/10.1007/978-3-030-34212-8_17

Zou, B., Liviero, S., Ma, Q., Zhang, W., Du, Y., & Xing, P. (2024). Exploring EFL learners' perceived promise and limitations of using an artificial intelligence speech evaluation system for speaking practice. *System*, *126*, 103497. https://doi.org/10.1016/j.system.2024.103497

Zou, B., Guan, X., Shao, Y., & Chen, P. (2023b). Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability*, *15*(4), 2872. https://doi.org/10.3390/su15042872

Chuxuan Xing is a Master candidate at The Department of English Language, Faculty of Languages & Linguistics, Universiti Malaya. Her area of research interest includes AI in second language learning.

Murad Abdu Saeed is a Senior Lecturer at The Department of English Language, Faculty of Languages & Linguistics, Universiti Malaya, Kuala Lumpur. His areas of research interest include Educational technology, feedback, engagement, feedback literacy, & collaborative writing. Orcid: https://orcid.org/0000-0003-2933-7929