Using AI Chatbots as Speaking Partners: Benefits, Limitations, and Learner Perceptions

Hashim Khan*., Irfan Hussain

Master's in Foreign Languages and Intercultural Communication, School of Foreign Languages, National Research University, Higher School of Economics, Moscow, The Russian Federation

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.100300254

Received: 08 March 2026; Accepted: 16 March 2026; Published: 03 April 2026

ABSTRACT

The integration of artificial intelligence chatbots as conversational partners in second language speaking instruction has gained considerable attention in recent years, driven by advances in natural language processing and automatic speech recognition technologies. This systematic review examines the pedagogical affordances, technical and instructional limitations, and learner perceptions of AI chatbots for speaking practice in English as a Foreign Language contexts. Drawing on 30 empirical studies and systematic reviews published between 2018 and 2025, this paper synthesizes evidence regarding the benefits of chatbot-mediated speaking practice, including increased practice opportunities, reduced speaking anxiety, enhanced learner engagement, and personalized learning experiences. However, the review also identifies persistent technical constraints---particularly speech recognition errors, limited contextual coherence, and inadequate pragmatic competence---alongside pedagogical challenges such as variable feedback quality, curriculum misalignment, and insufficient teacher preparation. Learner perceptions reveal a complex landscape: while students appreciate the accessibility, convenience, and low-pressure environment chatbots provide, they express frustration with unnatural interactions, repetitive responses, and limited corrective feedback. The paper situates these findings within established second language acquisition frameworks, including the Interaction Hypothesis, Affective Filter Hypothesis, and Sociocultural Theory, and discusses implications for pedagogical practice, teacher professional development, and future research directions. This review concludes that AI chatbots represent a valuable supplementary tool for speaking practice when thoughtfully integrated into curricula, but cannot replace human interaction in developing comprehensive communicative competence.

Keywords: AI chatbots, speaking practice, language learning, learner perceptions, second language acquisition, conversational AI

INTRODUCTION

The development of speaking proficiency remains one of the most challenging aspects of second language acquisition, particularly in foreign language contexts where opportunities for authentic oral interaction are limited. Traditional classroom instruction often fails to provide sufficient individualized speaking practice due to large class sizes, limited contact hours, and learners' reluctance to speak in front of peers. These constraints have motivated educators and researchers to explore technology-enhanced solutions that can extend speaking practice beyond the classroom and provide learners with low-stakes opportunities for oral production.

Recent advances in artificial intelligence, particularly in natural language processing and automatic speech recognition, have enabled the development of conversational AI systems capable of engaging in spoken dialogue with language learners. AI-powered chatbots---ranging from purpose-built language learning applications like ELSA Speak and Duolingo Max to general-purpose large language models such as ChatGPT with voice capabilities---now offer learners on-demand conversational partners accessible via smartphones and

computers. These systems promise to address longstanding barriers to speaking practice by providing unlimited, personalized, and anxiety-free opportunities for oral interaction.

The proliferation of AI chatbots in language education has generated substantial research interest, with numerous studies investigating their effectiveness, usability, and impact on learner outcomes. However, as Safitri et al. note, the existing research remains fragmented and lacks systematic synthesis, creating uncertainty for educators and policymakers regarding the educational value of chatbots in English language teaching. Moreover, while enthusiasm for AI-enhanced language learning is evident, critical questions persist regarding the quality of chatbot interactions, the pedagogical appropriateness of automated feedback, and the extent to which chatbot practice translates into improved communicative competence.

This paper addresses these gaps through a systematic review of recent empirical research on AI chatbots as speaking partners. Specifically, it examines: the pedagogical benefits chatbots offer for speaking practice, the technical and pedagogical limitations that constrain their effectiveness, and learner perceptions and attitudes toward chatbot-mediated speaking practice. By synthesizing evidence from 30 studies published between 2018 and 2025, this review provides a comprehensive assessment of the current state of knowledge and identifies priorities for future research and practice. The findings contribute to ongoing debates about the role of AI in language education and offer practical guidance for educators seeking to integrate chatbots into speaking curricula.

LITERATURE REVIEW

Theoretical Frameworks for AI-Mediated Speaking Practice

The use of AI chatbots as speaking partners can be understood through several established second language acquisition theories, though empirical studies have inconsistently invoked explicit theoretical frameworks. Three theories are particularly relevant for conceptualizing chatbot-mediated speaking practice: the Interaction Hypothesis, the Affective Filter Hypothesis, and Sociocultural Theory.

The Interaction Hypothesis, proposed by Long, posits that conversational interaction facilitates language acquisition when learners negotiate meaning, receive comprehensible input, and produce modified output in response to communication breakdowns. AI chatbots theoretically provide opportunities for such interaction by engaging learners in turn-taking dialogue, responding to learner utterances, and potentially offering clarification requests or reformulations. However, the extent to which current chatbot systems genuinely support negotiation of meaning remains contested, as many chatbots struggle to maintain coherent multi-turn conversations or provide meaningful interactional feedback.

The Affective Filter Hypothesis, developed by Krashen, suggests that emotional factors---particularly anxiety, motivation, and self-confidence---significantly influence language acquisition, with high anxiety creating a "filter" that impedes input processing. This framework is highly relevant to chatbot-mediated practice, as multiple studies report that learners experience reduced speaking anxiety when interacting with AI partners compared to human interlocutors. The non-judgmental, private nature of chatbot interaction may lower the affective filter, enabling learners to take risks and produce more output without fear of negative evaluation.

Sociocultural Theory, grounded in Vygotsky's work, emphasizes that language learning occurs through social interaction and scaffolded support within the learner's Zone of Proximal Development. From this perspective, chatbots can function as mediating tools that scaffold speaking practice through adaptive prompts, vocabulary support, and graduated task complexity. However, critics argue that chatbots lack the sophisticated interactional competence and contextual awareness necessary for truly responsive scaffolding, potentially limiting their effectiveness as mediators of learning.

Despite the theoretical relevance of these frameworks, research indicates that chatbot studies often lack explicit theoretical grounding, making it difficult to build cumulative knowledge about how and why chatbots support speaking development. This theoretical fragmentation represents a significant gap in the literature and underscores the need for more theory-driven research designs.

Evolution of Chatbot Technology in Language Learning

The application of chatbot technology to language learning has evolved considerably over the past two decades, progressing from rule-based systems with limited conversational capabilities to sophisticated neural language models capable of generating contextually appropriate responses. Early chatbots relied on pattern-matching algorithms and pre-scripted responses, resulting in interactions that learners often found repetitive, off-topic, or nonsensical. These systems struggled particularly with speech recognition for non-native speakers and provided minimal pedagogical scaffolding.

The emergence of large language models such as GPT-3, GPT-4, and their derivatives has dramatically expanded chatbots' conversational capabilities. Modern AI chatbots can engage in more natural, contextually coherent dialogue, adapt to diverse topics, and generate personalized responses based on learner input. The integration of advanced ASR technology has also improved voice interaction, enabling learners to practice spoken language rather than relying solely on text-based chat. Purpose-built language learning chatbots like ELSA Speak, Duolingo Max, and Replika now incorporate features specifically designed for L2 learners, including pronunciation feedback, vocabulary support, and structured conversation scenarios.

Despite these technological advances, significant limitations persist. Speech recognition systems continue to struggle with non-native accents, leading to frequent misrecognition that disrupts conversational flow. Chatbots often fail to maintain long, contextually rich exchanges, producing responses that lack pragmatic appropriateness or cultural sensitivity. Moreover, the "black box" nature of neural language models raises concerns about content accuracy, with chatbots occasionally generating incorrect information or culturally inappropriate responses.

Empirical Evidence on Chatbot Effectiveness

Recent meta-analytic and systematic review evidence provides insight into the overall effectiveness of chatbots for language learning. Wang et al.'s meta-analysis of 28 studies found a positive overall effect of chatbot use on language learning performance, indicating moderate effectiveness compared to non-chatbot conditions. The analysis revealed that effectiveness varied based on educational level, language proficiency, interface design, and interaction capability, suggesting that chatbot success depends on careful alignment between learner characteristics, pedagogical design, and technological affordances.

However, systematic reviews consistently note methodological limitations in the existing research base. Safitri et al.'s review of 111 studies found that research remains fragmented, with inconsistent outcome measures and limited evidence of robust, generalizable proficiency gains. Similarly, Jeon et al.'s systematic review of speech-recognition chatbots identified gaps in rigorous experimental designs and called for more controlled studies with standardized speaking assessments. These reviews suggest that while chatbots show promise, stronger empirical evidence is needed to establish their effectiveness for developing speaking proficiency.

Empirical studies examining specific chatbot implementations reveal mixed results. Hou's case study of the Doubao voice chatbot found that self-directed practice significantly enhanced learners' speaking abilities and increased practice frequency, though technical limitations and interaction challenges hindered sustained engagement. Ding et al.'s mixed-methods study demonstrated that AI-powered conversation bots reduced speaking anxiety and improved L2 speaking skills, though learners expressed concerns about feedback quality. Tai et al. found that generative AI chatbots improved elementary EFL learners' speaking skills in both individual and paired interaction modes, suggesting potential for collaborative chatbot-mediated practice.

Research Gaps and Study Objectives

Despite growing research interest, several critical gaps remain in understanding AI chatbots as speaking partners. First, few studies explicitly connect chatbot design and implementation to established SLA theories, limiting theoretical development. Second, comparative research examining different chatbot platforms is scarce, making it difficult for educators to make informed selection decisions. Third, longitudinal studies

tracking sustained chatbot use and its impact on speaking development over time are lacking. Fourth, research on teacher preparation and pedagogical strategies for integrating chatbots into curricula remains underdeveloped.

This systematic review addresses these gaps by synthesizing recent empirical evidence on chatbot benefits, limitations, and learner perceptions, with particular attention to pedagogical implications and practical recommendations for educators.

METHODOLOGY

This study employed a systematic review approach to synthesize existing research on AI chatbots as speaking partners in language learning. The review followed established guidelines for systematic literature reviews in educational technology research, focusing on identifying, evaluating, and synthesizing empirical studies and systematic reviews published between 2018 and 2025.

Search Strategy and Data Sources

A comprehensive literature search was conducted using two major academic databases: Google Scholar and SciSpace. The search query focused on the intersection of AI chatbots, speaking practice, and language learning, using terms such as "AI chatbots," "conversational AI," "speaking partners," "speaking practice," "language learning," "EFL," and "learner perceptions." The search was limited to peer-reviewed journal articles, conference proceedings, and systematic reviews published in English between 2018 and 2025 to capture recent developments in AI technology and their application to language education.

Inclusion and Exclusion Criteria

Studies were included if they: focused on AI chatbots or conversational AI systems used for speaking practice in L2 contexts, reported empirical data on benefits, limitations, or learner perceptions, involved EFL or ESL learners at any proficiency level, and were published in peer-reviewed venues. Studies were excluded if they: focused solely on text-based chatbots without speaking components, examined chatbots for purposes other than language learning, were purely theoretical without empirical data, or lacked sufficient methodological detail.

Data Extraction and Synthesis

The initial search yielded over 15,000 potential results across both databases. Following screening based on title and abstract relevance, 111 papers were identified for detailed review. From this pool, 30 studies representing the most relevant and methodologically rigorous research were selected for in-depth analysis. Data extraction focused on: study design and methodology, chatbot platforms and features examined, reported benefits of chatbot use, technical and pedagogical limitations, learner perceptions and attitudes, and theoretical frameworks employed.

Data synthesis employed a thematic analysis approach, identifying recurring patterns across studies related to benefits, limitations, and learner experiences. Findings were organized into three main categories: pedagogical benefits, technical and pedagogical limitations, and learner perceptions. Within each category, specific themes were identified and supported with evidence from multiple studies to ensure robustness of conclusions.

Quality Assessment

Study quality was assessed based on methodological rigor, sample size, clarity of research design, appropriateness of data collection and analysis methods, and transparency in reporting limitations. While formal quality scoring was not employed, preference was given to studies with larger samples, mixed-methods designs, and clear connections between research questions and findings. Systematic reviews and meta-analyses were weighted heavily due to their synthesis of multiple primary studies.

RESULTS

Pedagogical Benefits of AI Chatbots for Speaking Practice

The reviewed literature consistently identifies several key pedagogical benefits of using AI chatbots as speaking partners, though the strength and consistency of evidence varies across benefit categories.

Increased Practice Opportunities and Accessibility

One of the most frequently cited benefits is that chatbots provide learners with unlimited, on-demand opportunities for speaking practice outside classroom hours. Unlike human conversation partners, chatbots are available 24/7, enabling learners to practice at their convenience without scheduling constraints. This accessibility is particularly valuable in foreign language contexts where authentic speaking opportunities are scarce. Hou's case study found that learners using the Doubao voice chatbot reported significantly increased practice frequency compared to their prior routines, with participants engaging in self-directed speaking sessions multiple times per week. Similarly, Brinegar's review noted that chatbots extend practice opportunities beyond scheduled classroom hours, supporting self-directed learning and flexible use.

Reduced Speaking Anxiety and Lowered Affective Filter

Multiple studies report that chatbot interaction reduces speaking anxiety compared to speaking with teachers or peers. The non-judgmental, private nature of chatbot conversations creates a low-pressure environment that encourages spontaneous speaking and risk-taking. Ding et al.'s mixed-methods study found significant reductions in self-reported speaking anxiety among learners who practiced with AI conversation bots, with participants noting that the absence of human evaluation made them more willing to experiment with language. Belda-Medina et al. similarly reported that learners valued chatbots for reducing language anxiety and providing endless repetitions without judgment. This anxiety reduction aligns with the Affective Filter Hypothesis, suggesting that chatbots may facilitate language acquisition by lowering emotional barriers to output production.

Enhanced Engagement and Motivation

Research indicates that chatbots can increase learner engagement and motivation for speaking practice, particularly in the initial stages of use. The novelty and interactivity of conversational AI often generate enthusiasm and curiosity among learners. DU et al.'s systematic review found that AI chatbots enhanced students' confidence, engagement, and motivation in English-speaking learning, with the chatbot learning approach speeding up the English learning process. Alrajhi's study reported that EFL students found the chatbot interest-provoking and motivation-enhancing, while also alleviating writing anxiety. However, several studies note that engagement may decline over time due to novelty effects and the limitations of chatbot interactions.

Personalized and Adaptive Learning Experiences

AI chatbots offer potential for personalized learning through adaptive responses, customizable settings, and individualized feedback. Learners can adjust conversation topics, difficulty levels, and interaction styles to match their interests and proficiency levels. Zhi's exploratory study found that EFL learners perceived ChatGPT as a valuable learning partner that provided personalized learning experiences and immediate feedback tailored to individual needs. Ramana et al. highlighted that AI-powered chatbots offer interactive, personalized, and accessible language learning, complementing traditional methods with flexibility and engagement. This personalization supports differentiated instruction and enables learners to focus on specific areas for improvement.

Immediate Feedback and Conversational Turns

Chatbots provide immediate responses to learner utterances, sustaining conversational momentum and offering instant feedback on language use. This immediacy contrasts with delayed feedback in traditional classroom settings and enables learners to make real-time adjustments to their speech. Nhan's study found that students

recognized AI tools' potential to provide immediate, reliable feedback that reinforced strengths and identified areas for improvement. However, the quality and pedagogical appropriateness of this feedback varies considerably across chatbot systems, as discussed in the limitations section below.

Technical Limitations of AI Chatbots

Despite their benefits, AI chatbots face persistent technical constraints that limit their effectiveness as speaking partners.

Speech Recognition Errors and Accent Sensitivity

Automatic speech recognition remains a critical bottleneck for voice-enabled chatbots, particularly when processing non-native speech. Multiple studies report frequent misrecognition of learner utterances, especially for speakers with strong accents or lower proficiency levels. Gallacher et al.'s study found that the speech-to-text function was "unforgiving for L2 pronunciation," leading to communication breakdowns and learner frustration. These recognition errors disrupt conversational flow, force learners to repeat utterances multiple times, and may discourage continued use. Bin et al.'s literature review identified limitations in accommodating diverse accents as a persistent technical challenge requiring future development of more inclusive speech recognition systems.

Limited Contextual Coherence and Topic Maintenance

Chatbots frequently struggle to maintain coherent, contextually appropriate conversations over multiple turns. Studies report that chatbot responses often seem off-topic, fail to reference prior conversational content, or lack logical connection to learner utterances. Belda-Medina et al. found that learners criticized chatbots for providing "meaningless answers" and repetitive responses that did not advance the conversation. Alrajhi's study noted the "lack of extended conversations" as a demotivating factor, with the chatbot unable to sustain in-depth dialogue on complex topics. This limitation constrains the development of discourse-level speaking skills and reduces the authenticity of chatbot interactions.

Inadequate Pragmatic and Multimodal Competence

Current chatbots lack the pragmatic awareness and multimodal cues essential for authentic communication. They cannot interpret or produce facial expressions, gestures, intonation patterns, or other paralinguistic features that convey meaning in human interaction. Gallacher et al. reported that students perceived chatbot interactions as "lacking facial/verbal/emotional cues," resulting in unnatural communication. This absence of pragmatic and multimodal competence limits chatbots' ability to support the development of sociolinguistic and strategic competence, key components of communicative competence.

Content Accuracy and Cultural Appropriateness

The generative nature of large language models introduces risks of inaccurate, irrelevant, or culturally inappropriate content. Studies document instances of chatbots providing incorrect information, culturally insensitive responses, or content that contradicts established facts. Hung's study found that Vietnamese students expressed concerns about the accuracy and dependability of ChatGPT's language outputs, noting instances of wrong or culturally insensitive responses. Gutiérrez highlighted concerns about cultural biases in AI application, calling for enhanced digital competences for ethical AI utilization. These accuracy and appropriateness issues necessitate teacher mediation and raise questions about chatbots' reliability as autonomous learning tools.

Pedagogical Limitations of AI Chatbots

Beyond technical constraints, chatbots face pedagogical limitations that affect their instructional value.

Variable and Insufficient Corrective Feedback

A recurring concern is the quality and pedagogical appropriateness of chatbot feedback on learner errors. Many chatbots provide minimal or no explicit error correction, potentially allowing learners to practice inaccurate forms without remediation. Belda-Medina et al. found that chatbots "do not correct errors," which participants identified as counterproductive, particularly for lower-level students who need explicit guidance. When chatbots do provide feedback, it is often inconsistent, lacks pedagogical grounding, or fails to explain the nature of errors. Pratiwi et al. noted technical limitations in accurately assessing pronunciation and contextual nuances. This feedback variability undermines chatbots' potential to support accurate language development.

Curriculum Misalignment and Lack of Pedagogical Integration

Studies indicate that generic chatbot interactions often fail to align with specific curriculum goals, learning objectives, or assessment criteria. Without careful pedagogical design, chatbot practice may not target the linguistic forms, functions, or discourse patterns emphasized in formal instruction. Brinegar's review emphasized the need for "carefully selecting chatbots aligned with learning objectives" and "thoughtful integration" to maximize educational value. The disconnect between chatbot activities and curricular content can result in practice that does not transfer to classroom performance or assessment tasks.

Teacher Readiness and Professional Development Needs

Research reveals that while teachers are generally open to adopting chatbots, they often lack the knowledge and skills necessary for effective integration. Teachers report uncertainty about how to select appropriate chatbot tools, design pedagogically sound chatbot-mediated tasks, and scaffold learner use to prevent overreliance. Brinegar noted "knowledge gaps and priorities for enhancements" among teachers, highlighting the need for professional development. Without adequate teacher preparation, chatbot implementation may be superficial or pedagogically ineffective.

Risk of Shallow Practice and Overreliance

Technical breakdowns and limited interactional depth can lead learners to engage in shallow, formulaic practice rather than developing genuine communicative competence. When chatbot conversations are repetitive or lack meaningful content, learners may rehearse surface-level routines without deepening their linguistic or pragmatic knowledge. Additionally, concerns exist about learners becoming overly dependent on AI assistance, potentially undermining the development of autonomous communication skills. Nhan's study found that students expressed concerns about "the potential for overreliance on AI-driven methods," suggesting the need for balanced integration with traditional pedagogical approaches.

Learner Perceptions and Attitudes

Learner perceptions of AI chatbots as speaking partners reveal a complex and nuanced landscape, with both positive and negative dimensions.

Positive Perceptions: Convenience, Confidence, and Accessibility

Learners consistently appreciate the convenience and accessibility chatbots provide for speaking practice. The ability to practice anytime, anywhere without scheduling human partners is highly valued. Multiple studies report that learners feel more confident and comfortable speaking with chatbots than with human interlocutors, particularly in the early stages of language learning. Hou's case study found that participants valued the chatbot's "accessible, low-pressure environment, which encouraged spontaneous practice." Lee et al. reported that learners found LLM-based chatbots "highly capable of understanding dialogues in context," rating them positively for usefulness in L2 learning. Poláková et al.'s survey revealed "positive perceptions of this additional learning tool" and its perceived usefulness in second language acquisition.

Negative Perceptions: Unnatural Interaction and Limited Feedback

Despite positive aspects, learners frequently express frustration with the quality of chatbot interactions. Common complaints include unnatural, robotic conversation flow, repetitive responses, and off-topic replies that disrupt communication. Belda-Medina et al. found that learners criticized "unnatural computer-generated voices," "repetitive responses," and "limited understanding." Gallacher et al. reported that students perceived chatbots as "a novelty rather than a legitimate language-learning tool, lacking the richness of human interaction." Learners also express dissatisfaction with the quality and specificity of feedback, noting that chatbots often fail to provide the detailed error correction and explanations they need.

Mixed Perceptions: Usefulness vs. Limitations

Many learners hold simultaneously positive and critical views of chatbots, recognizing their utility for certain purposes while acknowledging significant limitations. Nugroho et al. found that learners appreciated ChatGPT as a conversation partner for practicing target languages but expressed concerns about "potential inaccuracies, cultural bias, repetitive responses, and the lack of credible sources." Alrajhi's study revealed that while students found the chatbot supportive of L2 practice and motivation-enhancing, they also identified demotivating factors including "the lack of extended conversations" and "sporadic irrelevant responses." This ambivalence suggests that learners view chatbots as useful supplementary tools rather than comprehensive solutions for speaking development.

Impact of Proficiency Level and Individual Differences

Learner perceptions appear to vary based on proficiency level, prior technology experience, and individual learning preferences. Some studies suggest that higher-proficiency learners may be better equipped to work around chatbot limitations and extract value from interactions. Alrajhi found that "high-intermediate students found it more useful for L2 practice" compared to lower-proficiency learners. However, other research indicates that lower-proficiency learners may benefit more from the scaffolded, low-anxiety environment chatbots provide. Individual differences in technology acceptance, learning styles, and attitudes toward AI also influence perceptions and sustained use.

DISCUSSION

Pedagogical Implications: Chatbots as Supplementary Tools

The evidence synthesized in this review suggests that AI chatbots are best conceptualized as supplementary tools for speaking practice rather than replacements for human interaction or comprehensive speaking instruction. Chatbots offer clear value in extending practice opportunities, reducing anxiety, and providing accessible conversational partners, particularly for learners who lack access to native speakers or feel inhibited in classroom settings. However, their technical and pedagogical limitations---especially regarding speech recognition accuracy, contextual coherence, and feedback quality---constrain their effectiveness as standalone learning solutions.

From a pedagogical perspective, chatbots align most closely with the Affective Filter Hypothesis, demonstrating consistent benefits in reducing speaking anxiety and creating low-pressure practice environments. The evidence for chatbots supporting the Interaction Hypothesis is more mixed: while chatbots provide conversational turns and immediate responses, their limited ability to negotiate meaning, maintain coherent multi-turn dialogue, and provide contingent feedback undermines their potential for facilitating acquisition through interaction. Similarly, chatbots' capacity to function as effective mediating tools within a sociocultural framework is constrained by their lack of sophisticated scaffolding and contextual awareness.

These findings suggest that chatbot integration should be strategic and purposeful, with clear alignment between chatbot affordances and specific learning objectives. Chatbots appear most appropriate for: providing supplementary practice to reinforce classroom instruction, building learner confidence and reducing anxiety before human interaction, offering rehearsal opportunities for specific conversational scenarios or linguistic

structures, and supporting autonomous learning for motivated students with adequate digital literacy. Chatbots are less suitable for: developing complex pragmatic or sociolinguistic competence, providing primary instruction in speaking skills, replacing teacher feedback and assessment, or serving as the sole source of conversational practice.

The Teacher's Role in Chatbot-Mediated Learning

The literature consistently emphasizes that effective chatbot integration requires active teacher involvement in selection, task design, scaffolding, and mediation. Teachers play several critical roles in chatbot-mediated speaking practice:

Selection and Evaluation: Teachers must evaluate available chatbot platforms based on pedagogical criteria, including alignment with curriculum goals, appropriateness for learner proficiency levels, quality of speech recognition and feedback, and cultural sensitivity. This requires developing digital literacy and critical evaluation skills specific to AI tools.

Task Design and Scaffolding: Rather than allowing unstructured chatbot use, teachers should design specific tasks and conversation scenarios that target curricular objectives and provide appropriate scaffolding. This might include pre-teaching vocabulary, modeling conversation strategies, setting specific goals for chatbot sessions, and providing post-interaction reflection activities.

Mediation and Feedback: Teachers must complement chatbot interaction with human feedback that addresses errors, clarifies misunderstandings, and provides the nuanced correction and explanation chatbots cannot offer. This hybrid approach combines the accessibility and anxiety-reduction benefits of chatbots with the pedagogical expertise and interactional competence of human teachers.

Monitoring and Support: Teachers should monitor learner engagement with chatbots, identify technical or motivational barriers, and provide ongoing support to sustain productive use. This includes helping learners develop strategies for working around chatbot limitations and maintaining realistic expectations about chatbot capabilities.

Professional Development: The evidence reveals significant teacher knowledge gaps regarding AI chatbot integration. Effective implementation requires professional development that addresses: understanding chatbot capabilities and limitations, evaluating and selecting appropriate tools, designing pedagogically sound chatbot-mediated tasks, integrating chatbot practice with classroom instruction, and addressing ethical considerations including privacy, bias, and overreliance.

Overcoming Barriers to Effective Implementation

Several strategies emerge from the literature for overcoming the technical and pedagogical barriers identified in this review:

Addressing Technical Limitations: While individual educators cannot directly improve speech recognition or contextual coherence, they can mitigate these limitations through careful tool selection, learner preparation, and task design. Selecting chatbots with better ASR performance for non-native speech, preparing learners for potential recognition errors, and designing shorter, more focused conversation tasks can reduce frustration from technical breakdowns. Educators should also advocate for continued technological development that prioritizes non-native speaker needs.

Enhancing Feedback Quality: To address variable feedback quality, teachers can implement hybrid feedback models that combine chatbot interaction with human review. For example, learners might record chatbot conversations for teacher analysis, complete self-assessment rubrics after chatbot sessions, or participate in peer review of chatbot interactions. Teachers can also explicitly teach learners to critically evaluate chatbot feedback and seek clarification when responses seem inappropriate.

Ensuring Curriculum Alignment: Chatbot practice should be explicitly connected to curriculum goals through structured tasks, clear learning objectives, and assessment integration. Teachers might create conversation prompts aligned with unit themes, design role-play scenarios that target specific linguistic functions, or use chatbot practice as formative assessment to identify areas needing classroom instruction.

Promoting Sustainable Engagement: To combat novelty effects and declining engagement, teachers should vary chatbot activities, set progressive challenges, and help learners track their progress. Gamification elements, peer sharing of chatbot experiences, and integration of chatbot practice into graded assignments can sustain motivation. However, care must be taken to avoid overreliance by balancing chatbot practice with human interaction.

Addressing Ethical Considerations: Teachers must address privacy concerns, cultural bias, and content accuracy issues proactively. This includes selecting chatbots with transparent privacy policies, discussing potential biases with learners, fact-checking chatbot-generated content, and teaching critical digital literacy skills. Educators should also model ethical AI use and help learners develop responsible practices.

Theoretical Contributions and Future Research Directions

This review reveals that chatbot research remains theoretically fragmented, with inconsistent invocation of SLA frameworks. Future research should explicitly ground chatbot studies in established theories to enable cumulative knowledge building and theory testing. Specific research priorities include:

a. **Theory-Driven Experimental Studies:** Research should design studies that explicitly test theoretical predictions about how chatbots support language acquisition. For example, studies could investigate whether chatbot interaction genuinely facilitates negotiation of meaning, whether anxiety reduction translates into improved acquisition, or whether chatbots can provide effective scaffolding within learners' ZPD.
b. **Longitudinal Studies:** Most existing research examines short-term chatbot use, leaving questions about sustained engagement and long-term learning outcomes unanswered. Longitudinal studies tracking chatbot use over months or years, with multiple speaking assessments, would provide crucial evidence about whether chatbot practice produces durable proficiency gains.
c. **Comparative Platform Studies:** The lack of comparative research makes it difficult for educators to select among available chatbot options. Future studies should conduct head-to-head comparisons of major platforms using standardized tasks and outcome measures.
d. **Integration and Implementation Research:** More research is needed on how to effectively integrate chatbots into curricula, including optimal task designs, feedback models, and teacher professional development approaches. Implementation science frameworks could guide research on scaling chatbot use in diverse educational contexts.
e. **Equity and Access Studies:** Research should examine how factors such as digital access, socioeconomic status, and prior technology experience affect chatbot use and learning outcomes. Studies should also investigate how chatbots can be designed to be more inclusive of diverse accents, dialects, and cultural backgrounds.

CONCLUSION

This systematic review of 30 studies published between 2018 and 2025 provides a comprehensive assessment of AI chatbots as speaking partners in language learning. The evidence reveals that chatbots offer significant pedagogical benefits, including increased practice opportunities, reduced speaking anxiety, enhanced engagement, and personalized learning experiences. These affordances make chatbots valuable supplementary tools for extending speaking practice beyond the classroom and supporting learners who lack access to human conversation partners.

However, persistent technical limitations---particularly speech recognition errors, limited contextual coherence, and inadequate pragmatic competence---constrain chatbots' effectiveness. Pedagogical limitations, including variable feedback quality, curriculum misalignment, and insufficient teacher preparation, further limit their

instructional value. Learner perceptions reflect this complexity: while students appreciate chatbots' convenience and anxiety-reducing qualities, they express frustration with unnatural interactions and limited feedback.

The findings suggest that chatbots are best positioned as supplementary tools that complement, rather than replace, human interaction and teacher-led instruction. Effective integration requires strategic selection, purposeful task design, teacher mediation, and ongoing professional development. When thoughtfully implemented within a broader speaking curriculum, chatbots can provide valuable low-stakes practice opportunities that build learner confidence and extend speaking time.

Future research should address current gaps through theory-driven experimental studies, longitudinal investigations of sustained use and learning outcomes, comparative platform evaluations, and implementation research examining effective integration strategies. As AI technology continues to advance, ongoing critical evaluation of chatbots' pedagogical value and limitations will be essential for ensuring that these tools genuinely support language learning goals rather than simply offering technological novelty.

Ultimately, the question is not whether AI chatbots can replace human interaction in speaking instruction---the evidence clearly indicates they cannot---but rather how they can be strategically deployed to address specific barriers to speaking practice while maintaining the centrality of human communication in language education. By recognizing both the affordances and constraints of current chatbot technology, educators can make informed decisions about when, how, and for whom chatbot-mediated speaking practice represents a valuable addition to the language learning toolkit.

REFERENCES

1. Safitri, N., et al., The impact of AI chatbots on English language learners' speaking proficiency: A systematic review, Journal of Research on English and Language Learning, 2025.
2. Brinegar, K., Chatbots as a supplementary language learning tool: advantages, concerns, and implementation, International Journal of Education and Social Science Research, 2023.
3. Bibauw, S., et al., Conversational AI and Chatbots, 2024.
4. Jeon, J., et al., A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models, Interactive Learning Environments, 2023.
5. Pratiwi, D., et al., Speaking Practice using ChatGPT's Voice Conversation: A Review on Potentials and Concerns, Journal of Language Intelligence and Culture, vol. 6, no. 1, 2024.
6. Wan, Q., et al., Using Call Annie as a Generative Artificial Intelligence Speaking Partner for Language Learners, RELC Journal, 2024.
7. Belda-Medina, J., et al., Using Chatbots as AI Conversational Partners in Language Learning, Applied Sciences, vol. 12, no. 17, 2022.
8. Cislowska, M., et al., Integration of Chatbots in Additional Language Education: A Systematic Review, European Journal of Educational Research, vol. 13, no. 4, 2024.
9. Xiao, Y., et al., A Review of Promises and Challenges of AI-Based Chatbots in Language Education through the Lens of Learner Emotions, Heliyon, 2024.
10. Reves, L., et al., Designing with Purpose: Pedagogical Reflections for the Development of AI Speaking Tools, The FLTmag, 2025.
11. Syair, M., et al., Artificial Intelligence as an Assisted-Language Learning Tool: A Systematic Review of Its Efficacy in Enhancing Speaking Skills Among EFL Learners, 2024.
12. Detailed insights synthesis, Theoretical frameworks used, 2025.
13. Long, M. H., The role of the linguistic environment in second language acquisition, in Handbook of Second Language Acquisition, W. C. Ritchie and T. K. Bhatia, Eds. San Diego: Academic Press, 1996, pp. 413-468.
14. Alrajhi, A., Artificial intelligence pedagogical chatbots as L2 conversational agents, Cogent Education, 2024.
15. Ericsson, P., Experiences of speaking with conversational AI in language education, 2024.

16. Gutiérrez, M., Artificial Intelligence in Language Education: Navigating the Potential and Challenges of Chatbots and NLP, 2023.
17. Krashen, S. D., Principles and Practice in Second Language Acquisition. Oxford: Pergamon Press, 1982.
18. Ding, L., et al., Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: a mixed methods study, Humanities & Social Sciences Communications, 2025.