Ampersand 14 (2025) 100224

Contents lists available at ScienceDirect

Ampersand

journal homepage: www.elsevier.com/locate/amper

AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications

Watcharapol Wiboolyasarin [a,b], Kanokpan Wiboolyasarin [c], Phornrat Tiranant [d], Nattawut Jinowat [e,*], Poomipat Boonyakitanont [f]

[a] Research Institute for Languages and Cultures of Asia, Mahidol University, Thailand
[b] School of Foreign Languages and South Asian and Southeast Asian Studies, Yunnan Minzu University, China
[c] Faculty of Management Science, Chandrakasem Rajabhat University, Thailand
[d] Faculty of Arts, The University of Hong Kong, Pok Fu Lam, Hong Kong
[e] Demonstration School, Suan Sunandha Rajabhat University, Thailand
[f] Looloo Technology Co., Ltd., Bangkok, Thailand

ARTICLE INFO

Keywords:
AI-driven chatbots
Second language teaching
Pedagogy
Personalized learning
Real-time feedback
Cross-cultural competence

ABSTRACT

This systematic review explores the efficacy of AI-driven chatbots in second language (L2) acquisition and examines their pedagogical implications. AI-powered chatbots have gained prominence in language education by offering personalized, adaptive learning experiences and continuous practice opportunities. Synthesizing findings from 30 empirical studies published between 2020 and 2024, this review investigates the impact of chatbots use across core language skills, including speaking, writing, reading, listening, grammar, and vocabulary. Results indicate notable improvements in learners' productive and receptive skills, especially in speaking and writing, due to features such as real-time feedback, anxiety reduction, and increase practice opportunities. While reading and listening also showed positive outcomes, their development was often constrained by chatbots' limited capacity for deep contextual interaction. Despite their benefits, chatbots faced challenges in replicating human-like emotional engagement and nuanced communication. Pedagogically, chatbots are promising tools for fostering personalized, autonomous L2 learning; however, their optimal use may depend on thoughtful integration with human instruction to enhance affective support and scaffold complex tasks. Further research is needed to enhance emotional responsiveness and support complex language processing in chatbot design.

1. Introduction

Language education has undergone significant transformation in recent decades, largely driven by advances in educational technology. Among these innovations, the application of artificial intelligence (AI) in second language (L2) learning has emerged as a promising development (Dong et al., 2024; Jinowat et al., 2024; Tolstykh and Oshchepkova, 2024; Wiboolyasarin et al., 2024a, 2024b). As global educational institutions adopt technology to support diverse learner needs, AI-powered tools—particularly chatbots—have gained momentum as accessible and adaptable solutions in language education (Chen, 2024; Lin and Chang, 2020; Son et al., 2023). These AI-driven chatbots, also referred to as conversational agents or digital assistants, are designed to simulate human-like dialogue using natural language processing (NLP) and machine learning (Adamopoulou and Moussiades, 2020; Mariani et al., 2023).

By offering instant feedback, round-the-clock availability, and immersive practice opportunities, chatbots promote behavioral, cognitive, and emotional engagement in language learning (Heung and Chiu, 2025; Wang and Xue, 2024). In L2 education, they help address common pedagogical challenges—such as limited speaking opportunities, performance anxiety, and lack of individualized support—by delivering personalized, low-pressure environments for learners to practice their skills (Cooray et al., 2024; Fathi et al., 2024; Kim and Su, 2024; Wu, 2024). With the continuing advancement of AI, integrating chatbots into classroom and autonomous learning contexts provides new ways to enrich language instruction through adaptive, student-centered approaches.

* Corresponding author.
E-mail address: nattawut.ji@ssru.ac.th (N. Jinowat).

https://doi.org/10.1016/j.amper.2025.100224
Received 30 January 2025; Received in revised form 5 May 2025; Accepted 12 May 2025
Available online 14 May 2025

1.1. AI-driven chatbots in language education

AI-driven chatbots are intelligent systems capable of simulating meaningful human-like interactions using natural language input. By integrating algorithms that recognize and generate language, these tools support various learning functions, including vocabulary reinforcement, grammar correction, conversation practice, and writing feedback (Escalante et al., 2023; Lee et al., 2023; Silitonga et al., 2023). For example, Duong and Suppasetseree (2024) found that students developed greater grammatical accuracy through consistent chatbot interactions, which offered rule-based corrections and repeated exposure to target structures.

More sophisticated chatbots adapt their responses based on learner input, allowing for personalized support in developing communicative competence (Kadir et al., 2024; Polakova and Klimova, 2024; Yahaya et al., 2024). These features enable learners to engage in conversation without fear of embarrassment, contributing to increased confidence and reduced anxiety (Tai and Chen, 2023; Yuan, 2024; Zhang et al., 2024a). Moreover, their availability and scalability make them particularly suitable for diverse educational settings, from formal classrooms to mobile-assisted learning environments (Guo et al., 2022; Maya and Camacho, 2024).

AI chatbots vary in complexity—from simple, rule-based systems to advanced models capable of machine learning. In language learning, they have been used to support all four macro-skills (listening, speaking, reading, and writing), as well as grammar and vocabulary development (Behforouz and Ghaithi, 2024; Belda-Medina and Kokošková, 2023; Guo and Li, 2024; Jeon, 2024; Liu et al., 2022; Polyzi and Moussiades, 2023; Zhang and Huang, 2024). Learners engage in real-time dialogue for oral practice or submit written texts to receive automated suggestions for improvement. These applications have been particularly valuable for repetitive tasks, individual practice, and supplementary instruction.

1.2. Literature review

Early research on educational chatbots (prior to 2020) focused primarily on their design and technical feasibility (Fryer and Carpenter, 2006; Jia and Chen, 2008; Jia and Ruan, 2008; Kerly et al., 2007), with applications in vocabulary instruction and basic conversational practice. As AI capabilities expanded, studies increasingly examined their use in broader language learning contexts, including grammar instruction, writing support, and oral proficiency development (Ruan et al., 2019; Wang et al., 2024c; Winkler and Söllner, 2018).

In recent years, the rise of large language models and generative AI has led to a surge of interest in more dynamic and responsive chatbot systems (Mariani et al., 2023; McGrath et al., 2024; Naik et al., 2024). Research conducted between 2020 and 2024 shows that chatbots are particularly effective at enhancing speaking fluency, writing quality, and learner motivation. For example, Fathi et al. (2024) found that AI-mediated speaking tasks improved learners' willingness to communicate (WTC), while studies by Tran (2025) and Zhang and Huang (2024) highlighted benefits in vocabulary acquisition and grammar correction.

However, existing literature also identifies several limitations. Many chatbots struggle with understanding nuanced or culturally embedded language, leading to generic or repetitive feedback (Annamalai et al., 2023; Guo et al., 2023; Terblanche, 2024). Additionally, the lack of emotional sensitivity and limited contextual awareness can hinder learner engagement, especially in tasks requiring empathy, creativity, or deep interpretation (Escalante et al., 2023; Yin et al., 2024; Yuan and Liu, 2025).

While prior reviews have explored chatbots in general education or emerging technology (e.g., Labadze et al., 2023; Okonkwo and Ade-Ibijola, 2021; Wollny et al., 2021), few have synthesized empirical findings specifically related to L2 learning across all core skills. Moreover, most studies focus on single-skill improvement or technical performance rather than examining the pedagogical implications of chatbot integration.

Therefore, a comprehensive review is needed to map the current research landscape, evaluate the efficacy of chatbots across language domains, and identify principles for effective integration. This review aims to fill that gap by analyzing trends, strengths, limitations, and pedagogical applications of AI chatbots in L2 instruction.

1.3. The present study

While existing research has shed light on the technological capabilities of AI-driven chatbots, there is a clear need for more systematic evaluation of their pedagogical impact in L2 learning (Wiboolyasarin et al., 2024c). Many studies have focused on individual skills or short-term outcomes (Kemelbekova et al., 2024; Lin and Mubarok, 2021; Zhang and Huang, 2024), but few offers a holistic analysis that cuts across language modalities, learning environments, and instructional practices. Furthermore, questions remain about which chatbot features and conditions contribute most effectively to language acquisition.

This systematic review addresses these gaps by synthesizing findings from 30 empirical studies on AI-driven chatbots in L2 education, focusing on two central research questions:

- RQ1: How effective are AI-driven chatbots in improving L2 acquisition?
- RQ2: What are the pedagogical implications of incorporating AI-driven chatbots into L2 teaching?

By answering these questions, the review offers insights into the practical integration of chatbots in language education and highlights directions for future research. The findings are intended to support educators in designing chatbot-enhanced curricula and guide researchers in refining AI-assisted language learning models. As AI technology continues to evolve, understanding its role in facilitating personalized, scalable, and engaging L2 learning is both timely and essential.

2. Method

2.1. Search strategy, criteria, and approach

This systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Page et al., 2021) to ensure transparency and rigor in study selection. The search was conducted in the Scopus database, which was chosen due to its comprehensive coverage of high-quality, peer-reviewed journals across disciplines, including education, linguistics, and computer science (Harzing and Alakangas, 2016; Zhang and Eichmann-Kalwara, 2019). Scopus provides broad indexing of international academic literature from major publishers, making it suitable for capturing scholarly work on AI and language education.

The search was conducted using combinations of keywords such as *"AI chatbot"*, *"artificial intelligence chatbot"*, *"language learning"*, *"second language education"*, and *"L2 teaching"*. Boolean operators (e.g., AND, OR) were used to refine the results. The search was limited to English-language articles published between 2020 and 2024, aligning with the recent acceleration in the development and adoption of AI technologies in education.

The inclusion and exclusion criteria used for screening are summarized in Table 1.

2.2. Screening process and study selection

The screening and selection process adhered to the PRISMA 2020 framework (see Fig. 1). A total of 1091 records were initially identified through a structured search of the Scopus database. These records were then screened based on inclusion criteria (1–2) and exclusion criteria

Table 1
Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
1. Peer-reviewed articles published in English between 2020 and 2024.	1. Studies focused exclusively on non-AI chatbots (e.g., simple rule-based systems without AI integration).
2. Articles with abstracts and keywords referencing AI-driven chatbots and L2 learning or teaching.	2. Papers addressing non-educational applications of AI-driven chatbots.
3. Empirical studies reporting primary data, including learner assessments, surveys, interviews, etc.	3. Studies that do not involve L2 learners or do not report educational/ pedagogical outcomes.
4. Research clearly describes context, participants, methodology, and educational relevance.	4. Theoretical papers, opinion pieces, and literature reviews lacking empirical data.

(1–4), resulting in the removal of:

- 19 records published outside the specified time frame (2020–2024),
- 446 records that were other document types (e.g., editorials, book chapters, conference summaries),
- 12 records not published in English, and
- 284 records that did not contain the required keywords in the title or abstract.

Following the removal of 761 duplicate records, 330 unique studies remained for screening. Of these, 293 were excluded for not meeting inclusion criteria 3–4, primarily due to the absence of empirical data or a lack of focus on L2 learners and pedagogical outcomes.

This resulted in 37 studies being identified for full-text retrieval. Of those, 32 full-text articles were successfully assessed for eligibility. Two articles were excluded at this stage—one due to insufficient data transparency and the other due to a lack of methodological rigor. = .

Finally, 30 empirical studies met all inclusion criteria and were included in the systematic review for full analysis.

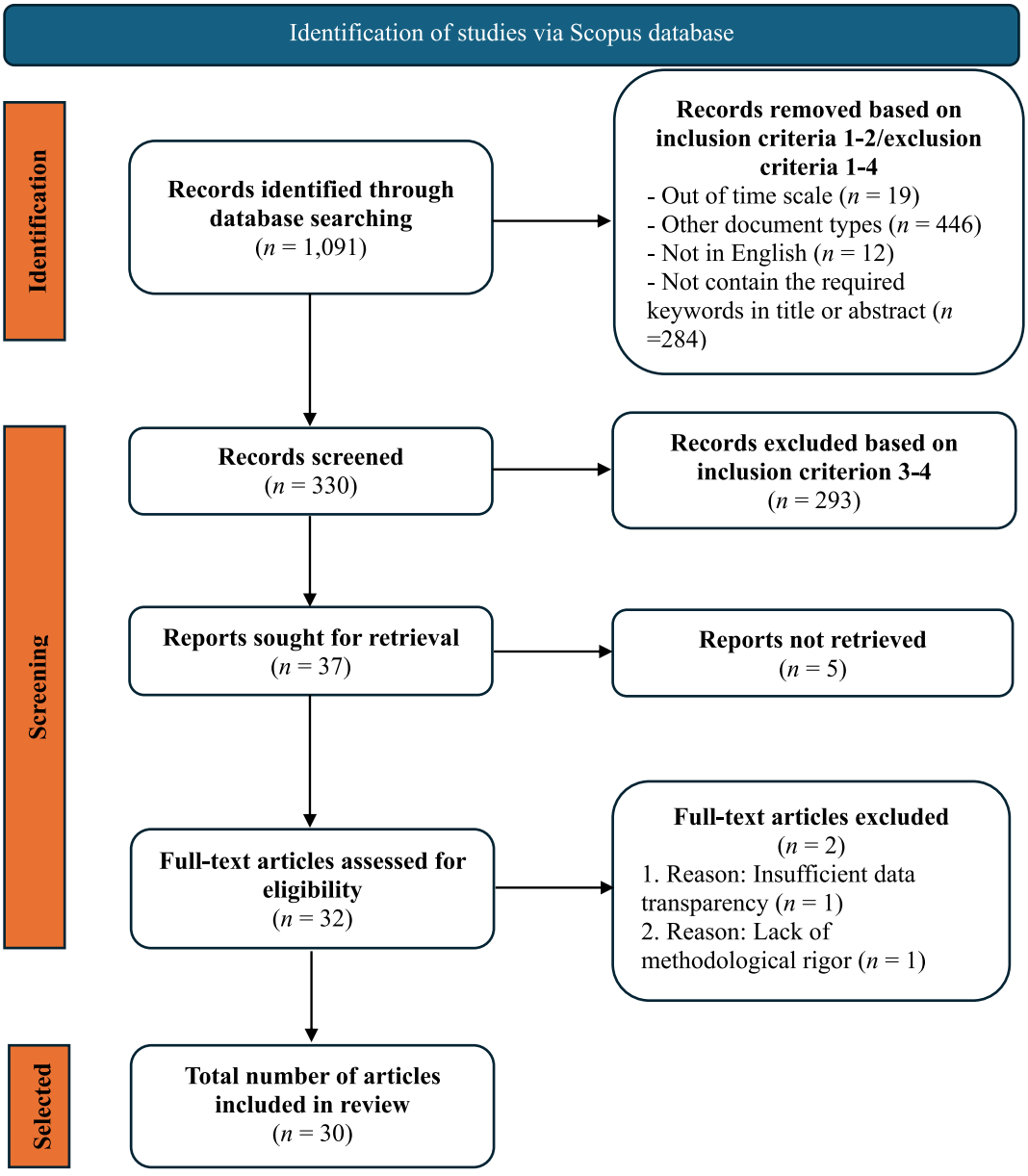

Fig. 1. The PRISMA flowchart for literature search and selection.

2.3. Study characteristics and coding approach

To facilitate analysis and comparison, the 30 studies included in this review were systematically classified according to a set of predefined characteristics. This categorization process was guided by research questions and informed by prior systematic reviews in the field of educational technology. Each study was reviewed in full and coded using a structured framework that included several key dimensions. These dimensions comprised the type of publication venue (such as education-focused journals versus interdisciplinary journals), the geographic region where the study was conducted (including Asia, Europe, North America, and Africa), and the educational level of the participants (categorized as primary, secondary, or post-secondary). Additionally, the studies were coded by research design, distinguishing between qualitative, quantitative, and mixed methods approaches. The learning context was also documented, capturing whether the intervention took place in classroom-based settings, through mobile-assisted learning, or as part of self-directed study. These variables were chosen to enable comparisons across diverse learning environments and to identify emerging patterns in the use of AI-driven chatbots for L2 acquisition. A summary of these study characteristics is presented in Appendix A.

2.4. Data extraction and quality review

To ensure consistency and comprehensiveness in data collection, a structured data extraction form was developed and applied to all included studies. Two researchers independently conducted the data extraction process to minimize potential bias and enhance the reliability of the review. Any discrepancies between the two reviewers were discussed and resolved through consensus. The extracted data captured essential study elements, including the research design (such as experimental, quasi-experimental, case study, or survey-based approaches), participant details (such as sample size, age group, education level, and language proficiency), and the type of chatbot used (e.g., text-based, speech-enabled, or multimodal AI).

Additional data included the duration and frequency of the intervention, the specific language skills targeted (speaking, writing, reading, listening, grammar, or vocabulary), and the reported learning outcomes, which ranged from language proficiency gains to changes in learner motivation or anxiety levels. Information about the instruments and measures employed in each study—such as pre- and post-tests, surveys, interview protocols, or interaction logs—was also recorded. Although a formal risk-of-bias assessment tool was not applied, the inclusion criteria ensured a minimum threshold of research quality by requiring clear methodological reporting and evidence-based findings. Nonetheless, future reviews could benefit from the use of standardized quality appraisal frameworks, such as the Mixed Methods Appraisal Tool (MMAT) or the Cochrane risk-of-bias checklist, to further enhance rigor.

3. Results

The 30 studies included in this systematic review reflect a surge in interest in AI-driven chatbots for language learning between 2020 and 2024, coinciding with the broader trend of integrating AI into educational contexts. These studies were geographically diverse, with research conducted in regions including Asia, Europe, North America, and Africa. A significant concentration of studies came from Asia, particularly countries such as China, Taiwan, and South Korea, highlighting the region's strong investment in digital language learning innovations.

The methodological designs employed across the studies were notably diverse. Quantitative research constituted approximately 43 % of the sample, while qualitative studies accounted for 7 %. Mixed-methods approaches were the most common, appearing in 50 % of the reviewed studies. This methodological variety underscores both the exploratory and evaluative nature of current research in chatbot-assisted L2 learning, as researchers aim to capture not only outcomes but also learner experiences and contextual factors.

Sample characteristics also varied. Most studies involved undergraduate students, with participant numbers ranging from small groups of five to larger samples exceeding 300 individuals. Some studies included elementary and secondary learners, and a few engaged mixed-age groups. This variation demonstrates the adaptability of chatbot technologies across different educational levels and age ranges.

In terms of targeted skills, speaking was the most commonly addressed area, with 57 % of the studies focusing on oral language development. Writing followed at 27 %, and reading was targeted in 17 % of the studies. Other studies focused on supporting grammar, vocabulary, or translation abilities. This distribution reflects the perceived potential of chatbots to enhance interactive and productive skills, particularly in speaking.

The chatbots used across the studies ranged from off-the-shelf applications, such as Replika and Mondly, to customized systems developed for specific pedagogical contexts. Intervention durations varied considerably, from single-session evaluations to longitudinal studies that spanned several weeks or months, indicating flexibility in how chatbot tools are integrated into instructional settings.

Regionally, the studies revealed different areas of emphasis. In Asia, research focused primarily on speaking, reading, and vocabulary development, often leveraging mobile and AI technologies to address common language learning challenges. In Europe, studies from countries such as the Czech Republic and Spain explored grammar instruction and learner interaction, often using mixed methods designs. In the Middle East and Africa, research from Egypt, Iran, and Oman centered on improvements in speaking and reading, while North American studies—mainly from the United States—examined chatbot use for speaking and translation skills.

Several common challenges emerged across the reviewed literature. While many studies reported enhanced language skills and learner engagement, some highlighted increased anxiety among participants, pointing to a need for chatbot designs that better support emotional well-being. Technical limitations, particularly in speech recognition and contextual understanding, also posed barriers to effective chatbot use. Despite these challenges, learner perceptions were generally positive, with most students appreciating the accessibility, feedback, and opportunities for practice provided by chatbots—although concerns about the "robotic" or impersonal nature of some interactions were noted.

3.1. Effectiveness of AI-driven chatbots in L2 acquisition (RQ1)

The results of 30 empirical studies were synthesized to evaluate the overall effectiveness of AI-driven chatbots in supporting L2 acquisition. The findings are structured by language skill areas to illustrate how chatbots contribute to improvements in speaking, writing, reading, listening, grammar, and vocabulary. Collectively, these studies demonstrate that AI chatbots have a substantial positive impact on L2 learning outcomes, including increased language proficiency, learner motivation, reduced anxiety, and enhanced engagement. Detailed findings are provided in a supplementary table (Appendix A) for interested readers.

3.1.1. Speaking skills

3.1.1.1. Enhancement of speaking performance. A total of 17 studies focused on the development of speaking skills through AI-driven chatbot interactions. The findings consistently show that chatbots significantly improve learners' oral proficiency, especially in structured dialogue tasks, pronunciation drills, and simulated conversations. For example, Lin and Mubarok (2021) integrated a mind map-guided AI chatbot (MM-AI) into a classroom setting where students rehearsed topic-based speaking tasks (e.g., describing routines, giving directions) using visual

mind maps. This structured rehearsal enhanced fluency, vocabulary retrieval, and conversational coherence.

Ye et al. (2022) implemented chatbot-mediated pronunciation and sentence construction drills, where students repeated model phrases and practiced forming questions and responses with real-time feedback. In other interventions, such as Yuan (2024) and Tai and Chen (2024), learners engaged in real-world speaking simulations—such as ordering food, interviewing peers, or navigating public transport—with the chatbot acting as a conversational partner. These tasks were integrated into blended learning modules or home-based speaking assignments, promoting spontaneity and vocabulary application. Fathi et al. (2024) included pair-based role-plays (e.g., doctor-patient, tourist-local) in speaking labs, where the chatbot facilitated turn-taking and provided immediate corrective feedback, boosting learners' fluency and expressive range.

3.1.1.2. Facilitation of practice in low-anxiety environments. Several studies emphasized the role of AI chatbots in creating supportive, non-judgmental speaking environments that reduce learner anxiety and promote WTC. In Kim and Su (2024), a self-paced chatbot module was developed for Korean learners, allowing students to rehearse speaking in private, low-pressure settings without the fear of peer evaluation. The activity included daily scenario-based interactions, such as introducing oneself or giving compliments, and showed a significant decrease in anxiety scores and increased speaking confidence.

Yuan (2024) described the use of chatbots for pre-class warm-up speaking sessions, where learners interacted with a chatbot to prepare for oral group discussions. This warm-up allowed students to practice key vocabulary and expressions, reducing their hesitation during live interaction. Yang et al. (2022) introduced informal chatbot chats outside of class, where students could initiate casual conversations on topics like music, hobbies, or weekend plans. These informal exchanges provided emotional safety, particularly for introverted learners.

However, Çakmak (2022) reported contrasting results. In this study, some students experienced heightened anxiety during chatbot conversations, especially when the system misinterpreted input or gave irrelevant responses. This outcome highlights the need to match chatbot complexity with learners' digital literacy and to provide onboarding or training to reduce initial apprehension.

3.1.1.3. Increased speaking opportunities and engagement. AI chatbots significantly increased the frequency and quality of speaking practice, enabling learners to engage in daily, meaningful oral interactions. Lin and Mubarok (2021) demonstrated how assigning short daily chatbot interactions (5–10 min) led to greater speech fluency over time. These sessions involved routine topics (e.g., daily schedules, personal interests) and required students to speak continuously, improving articulation and rhythm.

In Guo et al. (2023), chatbots like *Argumate* were used for academic debate simulations, where learners practiced expressing opinions, rebutting arguments, and using formal registers. These exercises were conducted as graded speaking tasks in EAP programs, with chatbots facilitating balanced dialogues and offering content feedback. Tai and Chen (2024) employed information gap activities, where students asked and answered questions during role-play scenarios (e.g., checking into a hotel), while Kemelbekova et al. (2024) used chatbots as rehearsal partners for oral presentations. Students used chatbots to test the clarity and flow of their speech, adapting their delivery based on the chatbot's follow-up prompts. The gamified nature of some platforms also maintained student interest through points, badges, and timed responses, sustaining motivation across multiple sessions.

3.1.1.4. Improved self-confidence and WTC. The reviewed studies consistently showed that chatbot-based speaking tasks enhanced learners' self-confidence and WTC, particularly through repetition, personalization, and progress tracking. In Ye et al. (2022), learners used the Microsoft Xiaoying chatbot to complete structured dialogue sets, such as restaurant conversations or making suggestions, with built-in pronunciation scoring. As their scores improved, students reported greater confidence in their speaking ability.

Yang et al. (2022) incorporated a dashboard feature showing speaking duration, turn counts, and feedback statistics, which encouraged goal setting and self-monitoring. Students reported that seeing measurable progress increased their willingness to participate in class discussions. Yuan (2024) found that students with low self-esteem became more talkative and expressive during non-evaluative chatbot sessions, often volunteering for oral tasks after practicing privately. Fathi et al. (2024) used partner-switching role-plays supported by chatbots to build gradual exposure, helping students take on both speaker and listener roles while building rapport and verbal agility.

3.1.1.5. Support for vocabulary development and conversational skills. AI chatbots promoted contextual vocabulary use and natural conversational flow through integrated vocabulary missions, quizzes, and real-time correction. Yang et al. (2022) developed task-based chatbot dialogues, where learners had to complete missions, such as planning a trip or solving a problem at school. Each task embedded new vocabulary (e.g., travel terms, school-related expressions), and the chatbot prompted learners to rephrase ideas, extend responses, or use synonyms.

Guo et al. (2023) assigned debate preparation exercises that required learners to study chatbot-generated vocabulary lists (e.g., ethical terms like "autonomy" or "bias") and use them in mock debates. This reinforced lexical application through discussion. Polakova and Klimova (2024) and Yuan (2024) supplemented their chatbot interactions with vocabulary quizzes, where learners received instant correction and contextual usage tips. Learners who practiced frequently demonstrated stronger word retention and were able to integrate new expressions into live conversations with more ease and flexibility.

3.1.1.6. Challenges in understanding and human-like interaction. Despite these gains, several studies reported limitations in chatbot understanding, emotional responsiveness, and natural language flow. Çakmak (2022) found that during open-ended interview simulations, learners were frustrated when the chatbot misinterpreted their input or responded with generic phrases, disrupting conversational rhythm and raising anxiety. Annamalai et al. (2023) echoed this in a storytelling task, where a rule-based chatbot failed to acknowledge tone or humor, making the interaction feel mechanical.

Belda-Medina and Kokošková (2023) evaluated chatbot use in mock job interviews and noted learner disengagement when the chatbot repeated pre-programmed responses or failed to follow up on nuanced answers. These shortcomings highlight the current limitations of rule-based or shallow NLP systems, especially in tasks requiring adaptive dialogue or emotional nuance. As a result, educators may need to complement chatbot activities with human-led feedback or peer discussion, especially in tasks involving rich narrative, negotiation, or social subtleties.

3.1.2. Writing skills

3.1.2.1. Collaborative writing and idea generation. Several studies highlighted how AI chatbots supported students in brainstorming, planning, and structuring arguments, especially during academic and persuasive writing tasks. In Guo et al. (2022), students used the chatbot *Argumate* as a virtual writing partner during in-class writing workshops. The chatbot posed analytical questions, suggested argumentative positions, and helped students organize claims and counterclaims for multi-paragraph essays. These chatbot-led sessions functioned as prewriting scaffolds, providing a collaborative interface to stimulate idea generation and logical reasoning.

Guo and Li (2024) extended this approach through project-based learning modules, where students built their own task-specific chatbots. These self-designed bots were integrated into essay planning sessions, assisting with thesis formulation, evidence grouping, and outlining subpoints. Similarly, Li et al. (2024a) employed *OpenBuddy* in online debate-style writing tutorials, where students argued with the chatbot on controversial topics. These interactions served as dialogic prewriting exercises, allowing learners to test their reasoning and anticipate counterarguments before drafting. Overall, chatbots were effectively embedded in classroom-based and blended learning environments to foster structured ideation and critical thinking.

3.1.2.2. Feedback, error correction, and peer review. AI chatbots were commonly used for automated writing feedback and to support peer review practices. In Guo et al. (2022), students revised drafts during asynchronous writing assignments using a combination of chatbots and grammar tools like Grammarly. The chatbot provided immediate feedback on sentence-level issues—grammar, coherence, and vocabulary—prompting students to revise content in iterative cycles. These revisions were part of digital writing portfolios maintained over the course of the semester.

In writing-intensive university courses, the chatbot *Eva* was integrated into a structured peer review system. During collaborative feedback workshops, *Eva* guided students through targeted review categories—such as clarity of thesis, argument organization, and grammar—helping them write more focused and actionable comments on peers' drafts. The chatbot also encouraged self-assessment by prompting students to compare their work to review guidelines. Similarly, Polakova and Klimova (2024) observed that in face-to-face classroom settings, chatbots helped learners identify transition errors and paragraph cohesion issues during timed essay writing, reinforcing linguistic accuracy and structural awareness.

3.1.2.3. Improvement in writing structure and logical coherence. Chatbots were shown to support improvements in essay structure, coherence, and logical flow, especially in multi-stage academic writing tasks. Guo et al. (2022) reported that students used chatbot-generated prompts during outline-to-draft activities, where they sequenced ideas, inserted appropriate transitions, and structured arguments across essay sections. These sessions were typically held in writing labs or tutorial blocks and included feedback checkpoints facilitated by the chatbot.

Guo and Li (2024) integrated chatbot-driven logic-mapping exercises into English for Academic Purposes (EAP) courses. Learners used the chatbot to identify weaknesses in argument chains and revise their claims based on the chatbot's diagnostic cues. However, Li et al. (2024a) cautioned that in more complex argumentative tasks—such as rebuttals or nuanced comparative essays—some chatbots (e.g., *OpenBuddy*) offered vague or overly formulaic suggestions, lacking the adaptability to handle subtle rhetorical shifts. These findings suggest that while chatbots can support coherence-building in structured tasks, more sophisticated designs are needed for higher-level academic writing.

3.1.2.4. Writing motivation and autonomy. Chatbot integration was positively associated with increased writing motivation and learner autonomy, particularly in self-paced or extra-curricular settings. Zhang et al. (2024b) found that students using chatbot-enhanced writing platforms reported higher levels of motivation due to the conversational and supportive tone of the chatbot, which mimicked a virtual writing coach. These platforms included interactive dashboards where students received real-time encouragement and topic suggestions.

In project-based modules described by Guo and Li (2024), students who developed custom chatbots for personalized writing support demonstrated enhanced self-regulation. These learners configured their bots to offer assistance on grammar patterns, vocabulary expansion, or paragraph organization—depending on their self-identified weaknesses. The chatbot served not only as a feedback provider but also as a personalized tutor, prompting revision and reflective writing. These activities were often carried out during homework assignments or online writing clinics, reinforcing independent learning habits.

3.1.2.5. Vocabulary development and language proficiency. Chatbots also contributed to enhanced lexical sophistication and grammatical accuracy, especially in academic and technical writing contexts. In Li et al. (2024a), *OpenBuddy* was used during academic writing tutorials, where it provided synonyms, sentence rewrites, and collocation suggestions to improve lexical diversity and sentence fluency in essays and literature reviews. Learners used this support to revise vague expressions or simplify overly complex structures.

Guo et al. (2023) observed chatbot use in collaborative translation and specialized content writing tasks, such as summaries of scientific articles or opinion editorials. Students interacted with the chatbot to confirm terminology, improve semantic precision, and use more discipline-specific vocabulary. Built-in dictionaries and multilingual support features enabled quick code-switching and clarification, fostering technical vocabulary acquisition. In classroom trials reported by Polakova and Klimova (2024), repeated engagement with chatbot-guided writing tasks correlated with improved vocabulary retention, as seen in pre- and post-test vocabulary assessments.

3.1.2.6. Challenges and limitations in chatbot-assisted writing. Despite these benefits, several limitations emerged regarding chatbot usability and linguistic competence in complex writing contexts. In Li et al. (2024a), students expressed dissatisfaction when chatbot feedback lacked specificity—particularly when addressing higher-order writing tasks like constructing counterarguments or synthesizing sources. Suggestions such as "revise your logic" or "make this clearer" were deemed too vague for meaningful revision.

Polakova and Klimova (2024) reported technical interruptions, such as delayed responses or interface crashes, during timed writing activities, which affected learner concentration. In Annamalai et al. (2023), reflective writing participants noted that chatbot feedback felt emotionally detached and failed to respond to nuances in tone or purpose. These limitations underscore the importance of combining chatbot-based scaffolding with human facilitation, particularly for tasks that demand nuanced interpretation or emotional sensitivity.

3.1.3. Reading skills

3.1.3.1. Self-regulated reading (SRR) and Chatbot support. An SRR is essential for developing L2 reading proficiency, and several studies demonstrated how AI chatbots fostered SRR through personalized, metacognitive support in digital reading environments. In Pan et al. (2024), the *ReadMate* chatbot was embedded within a mobile reading app designed for EFL learners. This tool enabled students to set reading goals, receive reading strategy tips, and track progress through periodic check-ins. Learners received personalized reading recommendations aligned with their language proficiency and interests, and were prompted with reflection questions and strategy suggestions, such as summarizing, skimming, or predicting content.

The chatbot was used primarily in self-paced, extracurricular reading programs, where learners interacted with the tool outside classroom hours using smartphones or tablets. This format gave students flexibility while promoting habit formation and autonomy, two critical aspects of sustained reading development. Learners reported that the chatbot's ongoing encouragement and adaptive support helped them persist with challenging texts, leading to better comprehension and greater motivation to read independently.

3.1.3.2. Engagement and interest through Chatbot interaction. AI chatbots also contributed to greater reading engagement and motivation,

especially when embedded in collaborative or gamified instructional contexts. In Liu et al. (2024), students participated in a learning-by-teaching activity, where they "trained" a chatbot to become a reading assistant. This involved inputting vocabulary explanations, paraphrased passages, and summaries from assigned readings to help improve the chatbot's comprehension accuracy. This competitive, goal-oriented approach turned reading into an interactive and social task, enhancing students' involvement and making the experience more dynamic.

Similarly, Liu et al. (2022) integrated chatbots into literature circle discussions, where the chatbot acted as a virtual peer. Students used the chatbot to summarize chapters, analyze character behavior, and ask interpretive questions, creating a dialogic reading environment that simulated group discussion. This peer-like interaction reduced the passive nature of individual reading and increased situational interest. Furthermore, the chatbot's nonjudgmental responses provided a safe space for experimentation and inquiry, especially valuable for reluctant readers who might hesitate to share in group settings.

3.1.3.3. Cognitive and linguistic scaffolding. Several studies highlighted the chatbot's role in providing on-demand scaffolding to support comprehension of complex texts. In Behforouz and Ghaithi (2024), students used a WhatsApp-based chatbot alongside their assigned readings. During these sessions, learners sent excerpts or vocabulary queries directly to the chatbot and received instant, contextualized support in return, including definitions, grammar explanations, and simplified rewritings. This real-time scaffolding enabled students to clarify meaning without disrupting the reading flow.

The activity was part of a blended learning model, where students completed reading assignments at home and brought their chatbot-enhanced understanding into classroom discussion. This approach supported differentiated instruction by giving learners the tools to manage comprehension challenges independently. Students who engaged with the chatbot outperformed those in the control group in both immediate and delayed comprehension assessments, particularly in navigating academic or dense texts, suggesting the chatbot served as a personalized support agent during reading tasks.

3.1.3.4. Reducing reading anxiety with chatbots. Foreign language reading anxiety (FLRA) is a persistent challenge in L2 acquisition, and chatbots were found to mitigate this affective barrier by creating private, low-stress reading environments. Zheng (2024) evaluated the GenAI-powered Reading Bot, used in tablet-based guided reading modules for Chinese secondary school students. These modules combined short reading passages with chatbot-supported comprehension tasks, such as rewording difficult expressions, answering open-ended questions, and receiving motivational messages.

The chatbot provided technological and emotional scaffolding, making learners feel more comfortable asking for help than they might in a teacher-led class. For instance, when students struggled with idiomatic expressions or unfamiliar grammar, the chatbot offered rephrased alternatives and calm encouragement, reducing fear of failure. This private, judgment-free interaction led to lower reported anxiety levels, greater persistence with difficult texts, and improved comprehension outcomes. The chatbot thus functioned not only as a language tool but also as an emotional support mechanism, reinforcing learners' confidence and willingness to engage with reading material.

3.1.4. Listening skills and translation

3.1.4.1. Enhancing listening comprehension through real-time interaction. Across the reviewed studies, AI-driven chatbots significantly enhanced listening comprehension by engaging learners in real-time, task-based listening and speaking activities. In Jeon (2024), students participated in simulated dialogue exercises, where they listened to chatbot-generated questions or prompts and responded orally in English. These interactive speaking tasks mimicked natural conversation flow and required learners to process spoken input in real time. The dialogues included common conversational expressions, idioms, and functional phrases, allowing learners to encounter language as it is used in authentic contexts.

During these tasks, the chatbot also modeled correct pronunciation and sentence structures, which learners were instructed to repeat and incorporate into dialogue completion exercises. Immediate oral feedback was provided, highlighting mispronunciations or misunderstood vocabulary, thus enabling learners to adjust their output and develop more accurate listening strategies. In addition, Jeon's study noted that role-play tasks were structured to include both formal and informal registers, helping students become familiar with varying tones and language styles—an essential skill for social and academic listening.

Other studies, such as Vanjani et al. (2020), implemented multilingual listening modules using chatbots that supported several target languages. Learners selected a language and engaged in listen-and-respond activities involving chatbot-generated audio outputs followed by comprehension checks or responses. These outputs ranged from casual conversations and instructions to short narratives, exposing students to different speech rates, regional accents, and registers. If learners struggled with understanding, the chatbot offered repeat, rephrase, or slower speech functions, which helped reinforce listening through multiple exposures. This kind of adaptive auditory scaffolding proved especially beneficial for learners with lower listening proficiency, as it supported individualized learning pace and comprehension monitoring.

The results of these studies indicated that integrating chatbots into role-play dialogues, pronunciation drills, and simulated listening tasks provided learners with continuous input and feedback cycles that closely mirrored human conversation. These activities enhanced learners' ability to interpret spoken language, manage listening anxiety, and develop both bottom-up (phonological decoding) and top-down (contextual inference) listening skills in an engaging and learner-centered environment.

3.1.4.2. Facilitating multilingual translation and cross-linguistic understanding. Several studies also explored how chatbots were integrated into translation and contrastive language learning tasks to promote multilingual literacy and intercultural awareness. In Vanjani et al. (2020), learners completed structured translation exercises by inputting sentences or short paragraphs in their first language (L1) and receiving real-time translations into a target language (L2 or L3), including Spanish, German, and Korean. These activities were embedded in scenario-based tasks, such as booking accommodation, ordering meals, or introducing oneself in a business setting, making the translations immediately relevant to real-world communication.

To deepen linguistic awareness, learners were often asked to compare the chatbot's translations to teacher-provided or peer-generated alternatives. These comparative analysis exercises encouraged students to reflect on differences in lexical choices, register, and grammatical structures, thereby developing their metalinguistic awareness. One particularly valuable design feature noted in the Korean-language chatbot was its ability to generate different responses based on formal or informal user inputs, serving as an entry point for instruction on sociolinguistic politeness strategies.

Learners also participated in peer discussion and reflection sessions where they reviewed and critiqued the chatbot's translations, pointing out overly casual phrases (e.g., translating "nice to meet you" as simply "pleasure") and learning to identify when adjustments were needed for more formal or culturally appropriate alternatives. These sessions fostered a critical stance toward AI-generated language, reinforcing that translation accuracy required both technological support and human judgment.

The integration of chatbots into context-driven translation tasks and multilingual listening activities helped learners move beyond surface-level language comparison to deeper cross-linguistic understanding. These dynamic, interactive exercises not only improved listening and translation proficiency but also encouraged learners to engage with cultural nuances and develop transferable skills in intercultural communication and language mediation.

3.1.5. Grammar and vocabulary

3.1.5.1. Real-time feedback and contextual learning. Grammar proficiency was a recurring focus in the reviewed studies, with chatbots supporting learners through interactive, task-based conversations embedded in authentic learning scenarios. In Annamalai et al. (2023), learners practiced grammar in the context of guided speaking and writing tasks, such as composing messages or holding simulated conversations, where the chatbot offered real-time correction and meta-linguistic explanations. These sessions were implemented in both classroom-based and homework settings, allowing for immediate feedback on articles, tenses, and sentence structure during natural language use.

Similarly, Belda-Medina and Kokošková (2023) employed the *Chatbot-Human Interaction Satisfaction Model (CHISM)* to assess grammar improvement using chatbots like *Andy* in structured grammar lessons. Activities included sentence transformation exercises, fill-in-the-blank grammar drills, and contextual grammar correction, where the chatbot responded with targeted feedback and explanations. Learners used these features during online practice modules and interactive grammar workshops, supporting independent learning with minimal instructor intervention.

In a blended learning context, Polakova and Klimova (2024) integrated chatbot interactions into weekly grammar reflection tasks, where students practiced forming complex sentences through scaffolded dialogues. The chatbot tracked errors and prompted learners to revise their input, creating a feedback loop that mimicked teacher guidance. These activities helped learners internalize grammar rules and apply them across multiple contexts, from casual chat to formal written responses.

These findings collectively suggest that chatbots were not merely grammar tutors, but interactive conversation partners embedded within meaningful communicative activities. Their greatest impact was observed when grammar correction was integrated into task-based learning settings, where learners needed to produce language in context and could receive timely, personalized corrections.

3.1.5.2. Personalized interaction and lexical growth. Vocabulary development was similarly supported through meaningful, task-oriented chatbot conversations that encouraged learners to use and retain new words in context. In Annamalai et al. (2023), students engaged in scenario-based dialogue practice, such as ordering food, booking travel, or discussing daily routines. The chatbot responded using context-relevant vocabulary and highlighted unfamiliar terms with embedded definitions. These activities were conducted in mobile learning environments and supported out-of-class self-study, giving learners autonomy over their pace and focus areas.

Belda-Medina and Kokošková (2023) conducted a comparative evaluation of four chatbots, identifying *Andy* as the most effective due to its adaptive vocabulary prompts tailored to learners' proficiency. Learners interacted with the chatbot through guided thematic tasks—such as discussing hobbies or social issues—where it provided synonyms, collocations, and example sentences that fit the topic. These conversations were integrated into formative speaking and writing exercises, allowing learners to expand and refine their lexical choices while staying on task.

In another study by Polakova and Klimova (2024), chatbot-based vocabulary review routines were incorporated into weekly classroom activities and extended into homework assignments. Students were asked to complete short storytelling or opinion tasks using target vocabulary, and the chatbot highlighted repetitive usage, suggested richer alternatives, and encouraged lexical diversity. Post-test assessments showed measurable gains in vocabulary range and contextual use, while learner reflections emphasized the benefit of interactive and repeated exposure to new words.

These vocabulary activities were not isolated drills; they were embedded in real communication tasks where learners negotiated meaning, responded to prompts, and used vocabulary with purpose. The chatbot's role extended beyond definition delivery—it served as a strategic learning partner, helping learners apply and personalize new language in context. This task-integrated design not only supported vocabulary retention but also made word learning relevant and motivating.

3.2. Pedagogical implications of AI-driven chatbots in L2 education (RQ2)

3.2.1. Personalized and adaptive learning support

AI-driven chatbots offer personalized learning experiences by adjusting task difficulty, feedback, and pacing based on learners' individual needs. In classroom and self-study settings, tools such as MM-AI, Andy, and CoolE Bot (Annamalai et al., 2023; Fathi et al., 2024; Polakova and Klimova, 2024; Yuan, 2024) provided targeted pronunciation drills, real-time vocabulary correction, and grammar coaching, allowing students to engage with content appropriate to their proficiency level. For example, in blended learning modules, learners used MM-AI to receive scaffolded sentence structures while preparing speaking tasks, while Andy delivered grammar explanations embedded within interactive dialogues. This adaptive feedback supported differentiated instruction, enabling teachers to assign chatbot-based tasks for remedial or advanced practice, particularly benefitting learners who struggle with traditional instruction or require more individualized attention.

3.2.2. Creating low-anxiety, non-judgmental learning environments

AI chatbots create emotionally safe practice spaces, especially important for learners with communication apprehension. Studies by Ebadi and Amini (2024) and Kim and Su (2024) documented the use of chatbot-mediated speaking warm-ups and asynchronous conversation modules, where students practiced everyday dialogues (e.g., greetings, restaurant orders) without peer observation. These activities promoted WTC by reducing fear of negative evaluation. Teachers can intentionally incorporate chatbot interactions into pre-task speaking routines or low-stakes oral assessments to lower performance anxiety and help learners build confidence before live class interactions.

3.2.3. Frequent opportunities for practice and increased engagement

Chatbots support high-frequency, low-stakes practice across all four language skills. As reported by Annamalai et al. (2023), Kemelbekova et al. (2024), and Tai and Chen (2024), learners completed daily speaking dialogues, chatbot-facilitated essay revisions, and reading comprehension checks, often in under 15 min. These bite-sized tasks were used as homework extensions, formative assessments, or in-class rotational stations, keeping learners actively engaged. Teachers can schedule routine chatbot tasks for homework, integrate them into flipped classroom models, or assign them during downtime to reinforce consistent language use and promote retention.

3.2.4. Support for grammar and vocabulary acquisition

AI chatbots serve as on-demand grammar tutors and vocabulary coaches, reinforcing linguistic accuracy through contextualized practice. In studies by Belda-Medina and Kokošková (2023) and Polakova and Klimova (2024), students used chatbots during academic writing labs or dialogue construction tasks, receiving real-time grammar

corrections and vocabulary substitutions. For instance, learners revising paragraphs with *OpenBuddy* received suggestions for improving cohesion or lexical variety, while those using Andy got explanations for subject-verb agreement errors within natural conversations. These capabilities make chatbots ideal for use in revision workshops, sentence-building activities, or targeted grammar drills aligned with curricular goals.

3.2.5. Balancing human interaction with AI assistance

While chatbots offer scalability and immediacy, they lack the nuanced interpretation and empathy provided by human interaction. Fathi et al. (2024) and Jeon (2024) demonstrated that students benefitted most when chatbots were paired with teacher-led debriefings or peer discussions. In Jeon's study, learners used a chatbot to complete listening dialogues and then discussed their answers in groups to clarify meaning. Teachers should balance chatbot use with guided reflection, error analysis sessions, or peer feedback, ensuring that learners receive both automated support and human judgment, particularly when addressing pragmatics, tone, or affective aspects of communication.

3.2.6. Addressing emotional engagement and response accuracy

Technical limitations—such as unnatural replies or misinterpretations—can frustrate learners, especially in tasks that require fluid, emotionally rich communication. Annamalai et al. (2023) and Belda--Medina and Kokošková (2023) found that students in storytelling or discussion-based tasks sometimes disengaged when chatbots gave irrelevant or generic responses. To mitigate these issues, teachers should prepare learners with strategy training, such as switching input modes (from speech to text), clarifying questions, or task redesigns that avoid ambiguous prompts. Educators might also use chatbot transcripts in class to teach error handling and promote critical reflection on machine-generated output.

3.2.7. Use of game-based learning to enhance motivation

Gamification was a key motivator in chatbot use, particularly when platforms included points, badges, ranks, or level progressions. Belda--Medina and Kokošková (2023) and Polakova and Klimova (2024) found that students using gamified chatbots practiced more frequently and with greater enthusiasm. For example, learners engaged in quiz-based vocabulary duels or completed chatbot missions that unlocked new language topics. Teachers can leverage these features by designing reward-based practice routines, offering class leaderboards, or using chatbots as part of review games to energize practice sessions, particularly for younger or demotivated students.

3.2.8. Promoting learner autonomy and self-regulated learning

Chatbots foster self-regulation by encouraging students to monitor their progress, set goals, and revise independently. In Annamalai et al. (2023) and Guo and Li (2024), learners built their own chatbots or used pre-programmed bots to generate feedback on writing drafts or speaking responses. This personalized loop encouraged revision, metacognitive reflection, and self-initiated improvement. Teachers can assign chatbot journaling tasks, self-paced review modules, or out-of-class assignments that emphasize learner control and initiative. These tasks reinforce habits of autonomous learning and reduce reliance on direct teacher correction.

3.2.9. Supporting flexible and ubiquitous learning environments

AI chatbots support ubiquitous learning, enabling language practice anytime and anywhere. Mobile-friendly platforms allowed students in Belda-Medina and Kokošková (2023) and Annamalai et al. (2023) to complete listening and writing tasks during commutes, lunch breaks, or evenings. These learners reported greater continuity in their learning routines. Teachers can promote this flexibility by encouraging daily chatbot check-ins, assigning just-in-time practice tasks, or embedding chatbot links in learning management systems (LMS). This strategy allows students to personalize when and where they engage with language material, supporting lifelong learning habits.

3.2.10. Enhancing social and cultural competence

Multilingual and culturally adaptive chatbots were especially effective in building intercultural competence. In Vanjani et al. (2020), learners used chatbots to practice formal vs. informal speech registers and explore cultural conventions across languages such as Spanish, Korean, and German. Tasks included role-plays (e.g., greeting elders vs. peers) and cross-cultural comparisons of phrases and tone. Teachers can integrate such chatbots into sociolinguistics modules, language-for-tourism courses, or intercultural communication units, helping learners not only improve linguistic fluency but also navigate social norms and cultural variation.

4. Discussion

This systematic review aimed to answer two primary research questions: (1) How effective are AI-driven chatbots in supporting L2 acquisition? and (2) What are the pedagogical implications of integrating chatbots into L2 instruction? To systematically address these questions and synthesize the wide range of findings, we adopted a three-pronged conceptual framework. This framework was inductively developed based on recurring themes identified in the reviewed studies, and it was deliberately designed to ensure alignment with both the analytical goals of the review and the research questions. The three dimensions—Human-AI Synergy, Skill-Specific Impact, and Pedagogical Considerations—were chosen to reflect the multifaceted nature of chatbot use in L2 education, allowing us to explore not only the overall effectiveness of chatbot interventions (RQ1) but also the practical considerations for implementation (RQ2).

Although only two research questions were formally articulated, we include the Skill-Specific Impact Dimension as a thematic sub-focus of RQ1. This addition enables a more nuanced examination of how chatbot use supports different language skills in varying instructional contexts.

The framework consists of the following three dimensions:

1. **Human-AI Synergy Dimension**: This dimension addresses RQ1 by examining the overall effectiveness of AI-driven chatbots in L2 acquisition. It focuses on how chatbots interact with or complement human instruction, identifying both the strengths and limitations of AI-mediated language learning environments. This analysis reflects broader pedagogical trends toward hybrid models that leverage both AI and human teaching (Belda-Medina and Kokošková, 2023; Annamalai et al., 2023).
2. **Skill-Specific Impact Dimension:** Also contributing to RQ1, this dimension provides a breakdown of chatbot effectiveness across key language domains: speaking, writing, reading, and listening. It highlights which skills show the most improvement from chatbot use and explores the instructional settings and tasks in which these gains are most apparent.
3. **Pedagogical Considerations Dimension**: This dimension addresses RQ2 by synthesizing the instructional implications of chatbot integration. It discusses how chatbots support personalized learning, engagement, learner autonomy, and emotional comfort, while also identifying limitations and offering strategies for classroom implementation.

This three-dimensional framework is an original contribution of this review and was not drawn from an existing theoretical model. Rather, it reflects a synthesized analytic approach grounded in the data and shaped by the scope of the review. By structuring the discussion in this way, we aim to provide both theoretical clarity and practical insights into how AI-driven chatbots can be effectively and responsibly integrated into L2 learning environments.

4.1. The Human-AI Synergy Dimension: effectiveness of AI-driven chatbots in improving L2 acquisition

This review highlights the substantial potential of AI-driven chatbots as tools for enhancing L2 acquisition, particularly when examined through the lens of Human-AI Synergy. The studies reviewed consistently demonstrate that chatbots contribute to L2 development by offering immediate, personalized feedback and enabling repeated, low-stakes practice. These features promote learner engagement and motivation, two critical predictors of language learning success (Fathi et al., 2024; Liu and Reinders, 2025; Yuan, 2024; Yuan and Liu, 2025).

A key strength of chatbots lies in their ability to create interactive, learner-centered environments that allow for personalized and autonomous learning experiences. Chatbots adjust to learners' responses in real time and can deliver scaffolded support based on individual performance—an approach aligned with Vygotsky's (1978) sociocultural theory and the concept of the Zone of Proximal Development. In practice, this allows learners to engage in self-paced learning activities that promote confidence and gradual mastery (Annamalai et al., 2023; Guan et al., 2025; Kim and Su, 2024; Yin et al., 2021).

Moreover, several studies emphasized chatbots' unique role in reducing affective barriers, particularly anxiety, by creating non-judgmental practice spaces (Cao and Wen, 2025; Kim and Su, 2024; Yuan, 2024; Zheng, 2024). This echoes earlier findings in computer-assisted language learning (CALL) research, which suggest that anxiety-reducing environments—such as those afforded by computer-mediated communication—can positively influence learner participation and oral production (Reinders and White, 2016; Tabuchi et al., 2024; Wang et al., 2024c). Chatbots extend this benefit by simulating natural conversation and offering repetitive practice without social pressure, thus increasing learners' WTC, especially in speaking tasks.

However, despite their strengths, AI chatbots show limitations in handling higher-order aspects of language learning. While they perform well in correcting surface-level errors (e.g., grammar, vocabulary), they are less effective at providing nuanced feedback on discourse structure, pragmatics, and tone—elements that are often better addressed through human scaffolding (Belda-Medina and Kokošková, 2023; Tudino and Qin, 2024). Additionally, several studies report that learners felt frustrated or isolated when the chatbot's responses lacked emotional resonance or conversational depth (Annamalai et al., 2023; Çakmak, 2022; Jasin et al., 2023), reflecting a key limitation in current natural language generation systems.

These findings point toward the pedagogical value of blended learning models, where chatbots complement rather than replace human instruction. A hybrid approach can draw on the chatbot's strengths—scalability, immediate feedback, and adaptability—while relying on teachers to provide context-rich explanations, emotional support, and metacognitive guidance. This aligns with existing literature on human-AI collaboration in education, which argues that optimal outcomes occur when technology is integrated in ways that augment (rather than automate) instruction (Holmes and Tuomi, 2022; Steiss et al., 2024).

The effectiveness of AI-driven chatbots in L2 acquisition is most fully realized when situated within a broader human-AI ecosystem. Chatbots offer scalable, accessible, and motivating tools for practice and feedback (Zhang and Huang, 2024), but their pedagogical impact is magnified when strategically integrated with teacher-mediated learning. Future chatbot design should continue to prioritize dialogue personalization and emotional responsiveness, while instructional models should emphasize the complementary roles of human educators and AI tools in language learning environments.

4.2. The Skill-Specific Impact Dimension: language skills benefiting most from Chatbot interactions

This review found that among the four core language skills, speaking and writing consistently demonstrated the most substantial improvements in chatbot-assisted learning environments. While reading and listening also showed positive outcomes, the gains were generally less pronounced. Understanding why these differences emerged is crucial for informing the pedagogical use of AI-driven chatbots in L2 education.

Speaking benefited most from chatbot integration due to the inherently dialogic and interactive nature of chatbot platforms. Unlike passive forms of practice, chatbot-mediated speaking tasks often simulate authentic conversation through structured role-plays, scenario-based dialogues, and turn-taking prompts (Huang et al., 2022; Wang et al., 2024b). These activities allow learners to repeatedly practice oral output in a low-pressure, judgment-free environment (Lin and Mubarok, 2021; Wang et al., 2024c; Ye et al., 2022). According to Kim and Su (2024), such environments are particularly valuable for learners with communication anxiety, as chatbots provide privacy, remove the fear of embarrassment, and offer instant corrective feedback—elements aligned with Krashen's (1982) affective filter hypothesis.

Furthermore, the immediate responsiveness of chatbots supports real-time fluency development, enabling learners to build automaticity in common communicative patterns (Tai and Chen, 2024). This echoes prior CALL research showing that oral fluency improves when learners are given frequent opportunities for output with immediate feedback (Ginkel et al., 2020). However, speaking gains are contingent upon the chatbot's ability to understand and respond intelligibly, and limitations in NLP may disrupt interaction flow (Çakmak, 2022).

Writing improvements were also notable, especially in contexts involving argumentation, peer review, and collaborative composition. Chatbots such as *OpenBuddy* and *Argumate* supported learners in brainstorming, organizing ideas, and receiving structural feedback during multi-paragraph essay writing tasks (Guo et al., 2022; Li et al., 2024a). This aligns with existing studies suggesting that AI tools enhance writing by scaffolding both content generation and revision (Khalifa and Albadawy, 2024; Tran, 2025). Moreover, the incorporation of chatbots into peer review workflows allowed students to give and receive structured feedback, which fostered metacognitive reflection and encouraged deeper engagement with writing mechanics (Polakova and Klimova, 2024).

That said, some limitations emerged. Chatbots struggled to handle nuanced aspects of writing—such as tone, argumentative coherence, or advanced rhetorical strategies—particularly in higher-level academic tasks. As noted by Li et al. (2024a), chatbot feedback on logic and organization was often vague, suggesting that AI remains more effective for micro-level corrections than macro-level discourse guidance.

In contrast, reading and listening showed more moderate gains, likely due to the nature of chatbot interactions, which prioritize output over input. In reading, gains were observed primarily when chatbots supported self-regulated reading (Pan et al., 2024, 2025) or provided real-time vocabulary and grammar scaffolding through mobile apps and messaging platforms (Behforouz and Ghaithi, 2024; Zhang and Huang, 2024). These tools enable learners to manage their reading progress, clarify confusion, and receive guidance during comprehension tasks. However, reading is typically a solitary, reflective process, and chatbots may offer less affordance for sustained, deep reading engagement unless carefully embedded into instructional routines.

Similarly, listening skills improved when chatbots delivered spoken input in varied registers or supported audio-based dialogues (Hong et al., 2016; Jeon, 2024; Vanjani et al., 2020). Yet chatbot speech synthesis and voice recognition technologies are not always sophisticated enough to replicate the full range of natural prosody or respond fluently to diverse accents and queries. As a result, the depth and quality of listening input may fall short of what learners encounter in authentic or multimedia-based listening tasks.

A noteworthy exception in this review was reported by Çakmak (2022), where learners showed improvement in speaking performance but experienced increased anxiety during chatbot interactions. This contrasts with most studies suggesting anxiety reduction and WTC improvement. The discrepancy may stem from contextual variables, such as technological unfamiliarity, learners' discomfort with robotic interaction, or frustration with chatbot misunderstandings, echoing concerns raised in affective computing literature (Picard, 2003). This highlights the importance of tailoring chatbot use to learner profiles and integrating sufficient orientation or training to reduce cognitive load during interactions.

The differential impact across skills can be attributed to how closely the chatbot's functionality aligns with the cognitive demands of each skill. Speaking and writing, which rely on interactive, generative output, are better served by the real-time, dialogic nature of chatbots. Reading and listening, which often require interpretive input processing, benefit less unless chatbots are explicitly designed to scaffold those tasks with interactive and adaptive input. These insights suggest that for chatbot integration to be most effective, instructional design must align chatbot tasks with the cognitive and social characteristics of each language skill, a recommendation echoed by Reinders and Benson (2017) in their call for more pedagogically grounded technology use in L2 instruction.

4.3. The Pedagogical Considerations Dimension: pedagogical implications of incorporating AI-driven chatbots in L2 teaching

The pedagogical integration of AI-driven chatbots into L2 instruction holds transformative potential, yet also presents new instructional demands. This dimension of the discussion explores how chatbots can support personalized learning, increase learner engagement, reinforce language knowledge, and promote learner autonomy—while also addressing practical challenges such as emotional engagement, teacher roles, and instructional design. Drawing on themes from the reviewed studies, this section offers actionable implications for L2 educators and curriculum designers.

4.3.1. Personalization and adaptive learning at scale

A key affordance of chatbots lies in their capacity for personalized and adaptive instruction. As shown in several studies (Fathi et al., 2024; Lin and Chang, 2023; Polakova and Klimova, 2024; Yuan, 2024), chatbots can tailor tasks and feedback to individual learners' proficiency levels, providing differentiated support in pronunciation, vocabulary usage, and sentence construction. This adaptability is particularly valuable in mixed-ability classrooms or self-paced learning environments, where traditional instruction may fall short in meeting every learner's needs.

Literature has long emphasized the value of adaptive systems in language learning (Heift and Schulze, 2007), and chatbots now bring this possibility into mainstream, scalable educational practice (Yekollu et al., 2024). When embedded thoughtfully into curricula, chatbots can help reduce instructional bottlenecks and provide continuous formative feedback—features aligned with principles of mastery learning and formative assessment (Black and Wiliam, 2009).

4.3.2. Addressing emotional engagement and human-like interaction

Despite their instructional advantages, chatbots face limitations in fostering affective connection and emotional engagement. Learners in several studies reported frustration or disengagement due to robotic or unresponsive chatbot behavior (Çakmak, 2022; Belda-Medina and Kokošková, 2023). Emotional disconnects were especially pronounced during tasks that required empathy, humor, or nuanced interaction, reinforcing concerns raised in prior work on digital limitations in social and emotional learning contexts (Picard, 2003; Li et al., 2024b).

To address this, we recommend a hybrid learning model that strategically combines chatbot use with human teacher intervention. Chatbots can efficiently manage repetitive, rule-governed tasks—such as grammar drills or vocabulary recycling—while teachers focus on high-touch areas like discourse-level feedback, cultural nuance, and emotional support. Such a model reflects broader shifts toward blended and "human-in-the-loop" AI pedagogies (Holmes and Tuomi, 2022), where the goal is not AI replacement but AI augmentation.

Additionally, educators should consider preparing students emotionally and technically for chatbot interaction. This may include short training sessions to clarify chatbot functionality and limitations, as well as scaffolding strategies to manage miscommunication. By setting realistic expectations and offering structured chatbot tasks (e.g., guided dialogues with error prompts), teachers can mitigate learner frustration and foster trust in the technology.

4.3.3. Redefining the Teacher's role and instructional integration

The adoption of chatbots also shifts the teacher's role from being a sole knowledge provider to learning facilitator and technology mediator. As learners begin to receive feedback from AI tools, teachers must shift focus toward curating effective chatbot tasks, interpreting chatbot outputs with students, and providing complementary human feedback where AI falls short (Almegren et al., 2024; Steiss et al., 2024). This redefined role is consistent with the evolving vision of teachers as orchestrators of digital learning environments (Beetham and Sharpe, 2013).

Educators should design activities that leverage the chatbot's strengths—such as real-time response, adaptive prompting, and availability outside class—while embedding them in meaningful communicative contexts. For instance, teachers might use chatbots as pre-writing brainstorming partners, conversation simulators for speaking warm-ups, or reading assistants for self-regulated reading plans (Hapsari and Wu, 2022; Shafiee Rad, 2025; Su et al., 2023). As shown in Liu et al. (2022) and Pan et al. (2024), when chatbot use is aligned with pedagogical goals and structured around authentic tasks, learners exhibit greater motivation, self-efficacy, and time-on-task.

4.3.4. Supporting grammar and vocabulary acquisition

Chatbots have proven effective in reinforcing grammar and vocabulary through low-stakes, contextualized repetition (Belda-Medina and Kokošková, 2023; Ghaithi and Behforouz, 2024; Polakova and Klimova, 2024; Tai and Chen, 2024). Unlike decontextualized drills, chatbot practice often involves embedded use of target structures in dialogue, thereby supporting incidental vocabulary learning and grammar-in-use.

Teachers can capitalize on this by assigning chatbot-based micro-tasks focused on target forms (e.g., practicing past tense in storytelling scenarios or vocabulary sets in shopping dialogues). Moreover, integrating chatbot logs or transcripts into classroom discussions can foster metalinguistic awareness and peer-corrective feedback, aligning with principles from sociocultural theory (Lantolf and Thorne, 2006).

4.3.5. Fostering autonomy and self-regulated learning

Chatbots also promote learner autonomy by offering on-demand support for L2 tasks, thereby encouraging self-directed learning habits (Wang et al., 2024a). As observed by Annamalai et al. (2023), learners used chatbots during out-of-class time to revisit concepts, practice speaking privately, or seek clarification. This aligns with Zimmerman's (2002) model of self-regulated learning, where access to timely feedback and self-monitoring tools fosters metacognitive control.

Educators can reinforce this by assigning chatbot tasks as part of homework routines, learning journals, or progress trackers. For example, learners could reflect on their chatbot interactions and submit summaries or error logs, prompting deeper engagement with both language and learning processes. When positioned as part of a larger reflective cycle, chatbots become more than tools for practice—they become scaffolds for lifelong learning habits.

5. Conclusion and future research directions

This systematic review highlights the growing relevance of AI-driven chatbots in L2 education, underscoring both their effectiveness in supporting skill development and their pedagogical implications. Across the reviewed studies, AI chatbots were found to be especially effective in promoting improvements in speaking and writing—two areas where immediate feedback, repeated practice, and low-pressure environments can make a substantial difference. These benefits were consistently linked to increased learner engagement, vocabulary development, grammatical accuracy, and overall confidence in language use.

Pedagogically, the findings suggest that educators can leverage chatbots to provide learners with additional conversational practice and writing feedback beyond classroom hours. Chatbots can function as accessible language partners, enabling self-paced and personalized learning, particularly for students who may be hesitant to speak in front of peers or who require more frequent input and correction. However, the review also reveals that chatbots are most effective when integrated within a broader instructional framework that includes human support, especially for handling emotionally nuanced feedback, discourse-level scaffolding, and cultural or pragmatic language use.

While the evidence supports the promise of chatbots in L2 education, challenges remain. Some learners reported dissatisfaction with chatbot limitations in human-like interaction, emotional responsiveness, and comprehension of complex input. To fully capitalize on the benefits of this technology, a hybrid learning model—where chatbots are used for skill-focused practice and human teachers provide higher-order guidance—emerges as a strategic approach for classroom integration.

Looking ahead, future research should address several gaps identified in the current literature. First, longitudinal studies are needed to evaluate the long-term effects of chatbot-assisted learning on language retention and proficiency. Second, there is a need to develop and test emotionally intelligent or culturally responsive chatbot systems, especially in under-researched contexts or languages beyond English. Investigating chatbot applications in under-represented languages and low-resource educational settings is particularly important to ensure equitable access and global applicability. Additionally, comparative studies across educational levels, proficiency bands, and learning environments (e.g., K-12 vs. higher education, formal vs. informal learning) would help determine how best to tailor chatbot use to diverse learner needs.

As AI continues to evolve, so too must our understanding of its pedagogical integration. This review calls for a balanced and intentional use of AI chatbots that complements, rather than replaces, the human dimensions of language learning.

6. Limitations

This review is subject to several limitations. First, publication bias may be present, as studies reporting non-significant or negative findings are less likely to be published in peer-reviewed journals and were therefore potentially underrepresented in this review. Second, the inclusion criteria were limited to studies published in English, which may have excluded valuable research from non-English-speaking communities—particularly in regions where chatbot integration in language education is growing rapidly. Third, there was considerable variability in chatbot design, technological capabilities, and instructional contexts across the reviewed studies, which may limit the generalizability of the findings to all language learning environments.

Additionally, although Scopus provided extensive access to relevant academic sources, it does not index all types of publications—such as gray literature or articles in non-indexed or regional journals. As a result, relying solely on Scopus may have led to the omission of some potentially relevant studies. To enhance the robustness of future research, scholars are encouraged to include studies from a broader range of databases and languages and to conduct more cross-contextual comparisons. Doing so will help ensure a more comprehensive understanding of how AI-driven chatbots function across diverse linguistic, cultural, and educational settings.

CRediT authorship contribution statement

Watcharapol Wiboolyasarin: Writing – review & editing, Writing – original draft, Validation, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation. **Kanokpan Wiboolyasarin:** Writing – original draft, Methodology, Investigation, Data curation. **Phornrat Tiranant:** Writing – review & editing, Visualization, Validation, Resources, Data curation. **Nattawut Jinowat:** Writing – review & editing, Writing – original draft, Validation, Resources, Methodology, Formal analysis. **Poomipat Boonyakitanont:** Software, Resources, Methodology.

Funding

This research project has been funded by Mahidol University (Fundamental Fund: fiscal year 2024 by National Science Research and Innovation Fund (NSRF)) (Grant No.: FF-181/2567).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Watcharapol Wiboolyasarin reports financial support was provided by Mahidol University. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Summary of characteristics of included studies

Article	Region	Sample	Level	Study design	Language skills targeted	Chatbot type	Duration	Language outcomes
Alrajhi (2024)	Saudi Arabia, Asia	143 students	Higher Education	Mixed methods (Questionnaire survey & Focus group discussions)	Writing	Tutor Mike (text-based pedagogical chatbot)	1 interaction session (25–30 min)	Positive experiences related to L2 practice, writing development, motivation, and alleviation of writing anxiety, though limitations in extended conversations and some irrelevant

(*continued on next page*)

(*continued*)

Article	Region	Sample	Level	Study design	Language skills targeted	Chatbot type	Duration	Language outcomes
								responses were reported.
Annamalai et al. (2023)	Malaysia, Asia	360 students	Higher Education	Mixed methods (survey and interviews)	Speaking, Writing, Grammar, Vocabulary	Duolingo, Mondly, Andy, Memrise, etc.	3 months	Chatbots were effective for language learning, particularly in practicing and facilitating conversations. However, social isolation was a challenge, with students perceiving chatbots as robotic and emotionless. Despite these issues, chatbots were deemed useful, especially during the COVID-19 pandemic.
Behforouz and Ghaithi (2024)	Oman, Asia	60 students	Higher Education	Quantitative (Pre-, Post-, Delayed Test)	Reading	WhatsApp Bot	3 weeks (10 sessions)	The experimental group outperformed the control group in post- and delayed post-tests, showing chatbots benefit reading skill development.
Belda-Medina and Kokošková (2023)	Spain, Czech Republic, Europe	237 students	Higher Education	Mixed methods (pre-post survey)	Speaking, Grammar, Interaction	Mondly, Andy, John Bot, Buddy.ai (App-Integrated Chatbots)	1-month intervention	Moderate satisfaction with chatbots; improvements in adapting to learner needs, multimedia integration, and speech technologies needed for a more human-like interaction.
Çakmak (2022)	Turkey, Europe	90 students	Higher Education	Mixed methods (Pretest-posttest)	Speaking, Anxiety management	Replika chatbot	12 weeks	Significant improvement in speaking performance with chatbots, but anxiety levels increased after the intervention; students had negative perceptions of chatbot interaction.
Ebadi and Amini (2024)	Iran, Asia	256 students	Higher Education	Mixed methods (SEM & thematic analysis)	Speaking	CSIEC chatbot	3 weeks	Social presence and human likeness positively impacted learner motivation and engagement. Students reported improved conversational skills and confidence.
Fathi et al. (2024)	Iran, Asia	65 students	Higher Education	Experiment	Speaking, WTC	Andy English Chatbot	12 weeks (20 min daily + 2 h weekly)	AI-mediated speaking activities were more effective in improving learners' speaking skills and WTC. Learners had positive attitudes towards AI-mediated instruction.
Guo and Li (2024)	Hong Kong, Asia	69 students	Higher Education	Mixed methods (Essays, questionnaires, reflections)	Writing	Self-made RAG chatbots via Poe	1.5-h workshop	Self-made chatbots improved writing motivation, led to clearer writing goals, increased confidence, and a

(*continued on next page*)

(*continued*)

Article	Region	Sample	Level	Study design	Language skills targeted	Chatbot type	Duration	Language outcomes
								more positive attitude towards writing. Students used chatbots for idea generation, grammar correction, and creating writing outlines.
Guo et al. (2023)	China, Asia	5 students	Higher Education	Qualitative study (Activity Theory)	Argumentative writing	Argumate (chatbot designed to scaffold argumentative writing)	One-time writing task	Students formed a close partnership with the chatbot, using various tools to facilitate collaboration; findings contribute to understanding student-chatbot collaboration in writing and offer insights for writing pedagogy.
Guo et al. (2023)	Hong Kong, Asia	24 students	Higher Education	Mixed Methods	Speaking (Debating)	Argumate (Argumentation chatbot)	90 min session	The chatbot inspired students in idea generation and reduced anxiety, but the uptake of chatbot-suggested ideas was limited in final debates. Students had a positive attitude towards chatbots, but some noted technical limitations.
Guo and Li (2024)	China, Asia	124 students	Higher Education	Quasi-experiment (Pretest-posttest)	Writing, Feedback	Eva (AI chatbot in EvaluMate platform)	3 weeks (9 peer review tasks)	AI-supported peer feedback significantly enhanced feedback quality and improved writing skills in content and language, but not in organization.
Jeon (2023)	South Korea, Asia	53 students	Elementary Education	Experiment	Vocabulary	Google Dialogflow chatbot	2 treatment sessions (25 min each)	CA-DA group had significantly higher vocabulary gains and diagnostic insights than CA-NDA and control groups.
Jeon (2024)	South Korea, Asia	36 students	Elementary Education	Exploratory qualitative study	Speaking and listening	Customized AI chatbots using Google's Dialogflow and Google Assistant	16 weeks	Chatbots provided interaction opportunities, immediate feedback, and reduced anxiety. Identified pedagogical, technological, and social affordances. Some constraints included speech recognition issues and lack of human interaction for some students.
Kemelbekova et al. (2024)	Kazakhstan, Asia	11 teachers, 51 students	Higher Education	Experiment	Speaking, Pronunciation	Text Cortex (vocab enrichment), Elsa Speak (voice recognition), Grammarly (grammar)	12 weeks, 3 months	Significant improvement in the experimental group's oral communication skills, with better intonation and stress patterns compared to the control group. No significant difference in pronunciation.

(*continued on next page*)

(*continued*)

Article	Region	Sample	Level	Study design	Language skills targeted	Chatbot type	Duration	Language outcomes
Kim and Su (2024)	China, Asia	65 students	Higher Education	Experiment (WTC questionnaire & semi-structured interview)	Speaking	Danbee AI chatbot (text & voice-enabled)	14 weeks (3 h per week)	Chatbot-mediated activities significantly improved WTC, reduced anxiety, and enhanced communication confidence in the experimental group compared to the control group.
Li et al. (2024a)	China, Asia	61 students	Higher Education	Quasi-experimental	Writing (Argumentative)	OpenBuddy (LLM-powered chatbot)	5 weeks	LLM-powered chatbot approach improved students' writing performance in content quality and language expression, provided personalized feedback, and enhanced motivation.
Lin and Mubarok (2021)	Taiwan, Asia	50 students	Higher Education	Experiment	Speaking	Replika app (speech-enabled)	5 weeks (100 min per week)	The mind map-guided AI chatbot approach promoted the students' English-speaking performances more than did the conventional AI chatbot approach.
Liu et al. (2022)	Taiwan, Asia	68 students	Elementary Education	Mixed-methods	Reading, Engagement	Custom AI chatbot (text-based)	6 weeks (80 min per week)	The chatbot increased students' social connection and helped maintain situational interest in reading. Students who interacted with the chatbot maintained a stable level of situational interest compared to the control group whose interest decreased.
Liu et al. (2024)	Taiwan, Asia	95 students	Elementary Education	Quasi-experimental	Reading	Teachable Q&A chatbot	9 weeks (50 min per week)	Chatbot training significantly enhanced students' reading interest and engagement; diverse question generation posed challenges, requiring deliberate training.
Pan et al. (2024)	Hong Kong, Asia	41 students	Higher Education	Mixed methods (Needs analysis, survey, interviews)	Reading, Self-regulation	Chatbot (ReadMate powered by ChatGPT 3.5)	2 months	ReadMate enhanced students' awareness and use of self-regulated reading strategies, but students faced challenges with topic alignment and assessing their own reading levels.
Polakova and Klimova (2024)	Czech Republic, Europe	58 students	Higher Education	Mixed methods (pre-test/post-test, questionnaire)	Vocabulary, Grammar, Writing, Speaking	Alex (Smart AI Chatbot)	4 weeks (3 sessions per week, 1000 characters/ session)	Significant improvement in vocabulary and grammar; positive student perceptions of chatbot usefulness, but technical issues and unfinished sentences reported.

(*continued on next page*)

(*continued*)

Article	Region	Sample	Level	Study design	Language skills targeted	Chatbot type	Duration	Language outcomes
Shazly (2021)	Egypt, Africa	48 undergraduate EFL learners (38 completed the study)	Higher Education	Quasi-experimental, pretest-posttest	Speaking, Anxiety management	Mondly and web chatbots	8 weeks (10 h intervention, 100 min per week)	AI-driven chatbots showed potential for improving speaking performance (mean score increased from 5.4 to 8.0), but anxiety levels slightly increased post-intervention (from 46 % to 51 % reporting anxiety).
Tai and Chen (2024)	Taiwan, Asia	85 students	Elementary Education	Mixed methods (Pre/post-test, Interviews)	Speaking	CoolE Bot (GAI chatbot with ASR)	3 weeks (5 days a week, 45 min per day)	Both individual and paired interactions with CoolE Bot significantly improved EFL speaking skills compared to the No-Bot group. No significant difference between I-Bot and P-Bot groups.
Vanjani et al. (2020)	USA, North America	49 students	Higher Education	Mixed methods (Transcripts evaluation)	Speaking, Translation	Tutor Mike (Multilingual chatbot using Google Translate)	1-time evaluation	Students found the chatbot's translations and responses generally good, with 6.08/7 average satisfaction, although some responses were out of context.
Yang et al. (2022)	South Korea, Asia	314 students (177 elementary students; 137 high school students)	Elementary and Secondary Education	Mixed methods (Conversation logs and questionnaires)	Speaking	Task-based AI chatbot "Ellie" developed by researchers	Varied (1–3 sessions over 3 weeks)	The chatbot assisted L2 learners in developing conversation; high task success rates (88.3 %); students had neutral to positive attitudes towards using the chatbot in EFL classes; suggestions for future improvements in chatbot design were discussed.
Ye et al. (2022)	China, Asia	50 students	Secondary Education	Experiment	Speaking	Microsoft Xiaoying	28 days	Grammar and pronunciation accuracy improved significantly; positive perception from participants.
Yuan and Liu (2025)	China, Asia	74 students	Elementary Education	Experiment	Oral English proficiency, WTC	Mondly (AI Chatbot)	3 months (10 min 3 times a week)	The AI chatbot group significantly improved in oral English proficiency and WTC compared to the control group. Chatbot helped reduce anxiety and boost fluency.
Zhang and Huang (2024)	USA, North America	52 students	Secondary Education	Quasi-experiment	Vocabulary (receptive and productive), Incidental learning	LLM-based chatbot	8 weeks	EG outperformed CG in both receptive and productive vocabulary, higher retention, and promoted incidental vocabulary learning.
Zhang et al. (2024b)	Hong Kong, Asia	30 students	Higher Education	Mixed methods (Pre-post tests, interviews)	Writing (Logical fallacies)	ManyChat (on Facebook Messenger)	5 weeks (autonomous learning)	Chatbot less effective in knowledge acquisition but more effective in improving motivation than

(*continued on next page*)

(*continued*)

Article	Region	Sample	Level	Study design	Language skills targeted	Chatbot type	Duration	Language outcomes
Zheng (2024)	China, Asia	84 students	Secondary Education	Mixed-methods quasi-experimental pre-test/post-test, qualitative interviews	Reading	GenAI-based "Reading Bot"	5 sessions (45 min each)	website-based learning. Significant reduction in FLRA in EG but no significant difference in FLRA and FLRP between EG and CG. Qualitative data indicated that the chatbot provided pedagogical, linguistic, cognitive, and affective affordances, but some challenges like information overload and occasional inaccuracies were noted.

References

Adamopoulou, E., Moussiades, L., 2020. An overview of chatbot technology [Paper Presentation]. In: 16th IFIP WG 12.5 International Conference on Artificial Intelligence Applications and Innovations (AIAI 2020), Neos Marmaras, Greece. https://doi.org/10.1007/978-3-030-49186-4_31.

Almegren, A., Mahdi, H.S., Hazaea, A.N., Ali, J.K., Almegren, R.M., 2024. Evaluating the quality of AI feedback: a comparative study of AI and human essay grading. Innovat. Educ. Teach. Int. 1–16. https://doi.org/10.1080/14703297.2024.2437122.

Alrajhi, A.S., 2024. Artificial intelligence pedagogical chatbots as L2 conversational agents. Cogent Educ. 11 (1), 1–18. https://doi.org/10.1080/2331186X.2024.2327789.

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. Comput. Educ. Artif. Intell. 5, 1–9. https://doi.org/10.1016/j.caeai.2023.100153.

Beetham, H., Sharpe, R., 2013. Rethinking Pedagogy for a Digital Age: Designing for 21st Century Learning, second ed. Routledge. https://doi.org/10.4324/9780203078952.

Behforouz, B., Ghaithi, A.A., 2024. Investigating the effect of an interactive educational chatbot on reading comprehension skills. Int. J. Eng. Pedagogy 14 (4), 139–154. https://doi.org/10.3991/ijep.v14i4.48461.

Belda-Medina, J., Kokošková, V., 2023. Integrating chatbots in education: insights from the chatbot-human interaction satisfaction model (CHISM). Int. J. Educat. Tech. High. Educat. 20, 1–20. https://doi.org/10.1186/s41239-023-00432-3.

Black, P., Wiliam, D., 2009. Developing the theory of formative assessment. Educ. Assess. Eval. Account. 21 (1), 5–31. https://doi.org/10.1007/s11092-008-9068-5.

Çakmak, F., 2022. Chatbot-human interaction and its effects on EFL students' L2 speaking performance and anxiety. Novitas-ROYAL (Res. Youth Lang.) 16 (2), 113–131.

Cao, B., Wen, C., 2025. Does greater chatbot usage lead to more positive affective attitudes toward AI? A moderated mediation model of relational capacity, competence, and privacy concerns in a two-wave panel study. Telematics Inf. 98, 1–13. https://doi.org/10.1016/j.tele.2025.102263.

Chen, M.-R.A., 2024. Metacognitive mastery: transformative learning in EFL through a generative AI chatbot fueled by metalinguistic guidance. Educ. Technol. Soc. 27 (3), 407–427. https://doi.org/10.30191/ETS.202407_27(3).TP05.

Cooray, S., Hettiarachchi, C., Nanayakkara, V., Matthies, D., Samaradivakara, Y., Nanayakkara, S., 2024. Kavy: fostering language speaking skills and self-confidence through conversational AI [Paper Presentation]. In: Augmented Humans International Conference, Melbourne, Australia. https://doi.org/10.1145/3652920.3652944.

Dong, W., Pan, D., Kim, S., 2024. Exploring the integration of IoT and Generative AI in English language education: smart tools for personalized learning experiences. J. Computat. Sci. 82, 1–9. https://doi.org/10.1016/j.jocs.2024.102397.

Duong, T., Suppasetseree, S., 2024. The effects of an artificial intelligence voice chatbot on improving Vietnamese undergraduate students' English speaking skills. Int. J. Learn. Teach. Educat. Res. 23 (3), 293–321. https://doi.org/10.26803/ijlter.23.3.15.

Ebadi, S., Amini, A., 2024. Examining the roles of social presence and human-likeness on Iranian EFL learners' motivation using artificial intelligence technology: a case of CSIEC chatbot. Interact. Learn. Environ. 32 (2), 655–673. https://doi.org/10.1080/10494820.2022.2096638.

Escalante, J., Pack, A., Barrett, A., 2023. AI-generated feedback on writing: insights into efficacy and ENL student preference. Int. J. Educat. Tech. High. Educat. 20, 1–20. https://doi.org/10.1186/s41239-023-00425-2.

Fathi, J., Rahimi, M., Derakhshan, A., 2024. Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. System 121, 1–17. https://doi.org/10.1016/j.system.2024.103254.

Fryer, Luke, Carpenter, Rollo, 2006. Bots as language learning tools. Lang. Learn. Technol. 3 (10), 8–14 https://doi.org/10125/44068.

Ghaithi, A.A., Behforouz, B., 2024. The use of an interactive chatbot in grammar learning. J. Educat. Online 21 (4), 1–10.

Ginkel, S.V., Ruiz, D., Mononen, A., Karaman, C., de Keijzer, A., Sitthiworachart, J., 2020. The impact of computer-mediated immediate feedback on developing oral presentation skills: an exploratory study in virtual reality. J. Comput. Assist. Learn. 36 (3), 412–422. https://doi.org/10.1111/jcal.12424.

Guan, R., Raković, M., Chen, G., Gašević, D., 2025. How educational chatbots support self-regulated learning? A systematic review of the literature. Educ. Inf. Technol. 30, 4493–4518. https://doi.org/10.1007/s10639-024-12881-y.

Guo, K., Li, D., 2024. Understanding EFL students' use of self-made AI chatbots as personalized writing assistance tools: a mixed methods study. System 124, 1–17. https://doi.org/10.1016/j.system.2024.103362.

Guo, K., Wang, J., Chu, S.K.W., 2022. Using chatbots to scaffold EFL students' argumentative writing. Assess. Writ. 54, 1–6. https://doi.org/10.1016/j.asw.2022.100666.

Guo, K., Zhong, Y., Li, D., Chu, S.K.W., 2023. Investigating students' engagement in chatbot-supported classroom debates. Interact. Learn. Environ. https://doi.org/10.1080/10494820.2023.2207181.

Hapsari, I.P., Wu, T.-T., 2022. AI chatbots learning model in English speaking skill: alleviating speaking anxiety, boosting enjoyment, and fostering critical thinking. In: Huang, Y.-M., Cheng, S.-C., Barroso, J., Sandnes, F.E. (Eds.), Innovative Technologies and Learning. Springer, pp. 444–453. https://doi.org/10.1007/978-3-031-15273-3_49.

Harzing, A.W., Alakangas, S., 2016. Google scholar, Scopus and the web of science: a longitudinal and cross-disciplinary comparison. Scientometrics 106 (2), 787–804. https://doi.org/10.1007/s11192-015-1798-9.

Heift, T., Schulze, M., 2007. Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues. Routledge. https://doi.org/10.4324/9780203012215.

Heung, Y.M., Chiu, T.K.F., 2025. How ChatGPT impacts student engagement from a systematic review and meta-analysis study. Comput. Educ. Artif. Intell. 8, 1–13. https://doi.org/10.1016/j.caeai.2025.100361.

Holmes, W., Tuomi, I., 2022. State of the art and practice in AI in education. Eur. J. Educ. 57 (4), 542–570. https://doi.org/10.1111/ejed.12533.

Hong, Z.-W., Huang, Y.-M., Hsu, M., Shen, W.-W., 2016. Authoring robot-assisted instructional materials for improving learning performance and motivation in EFL classrooms. J. Educat. Tech. Soc. 19 (1), 337–349.

Huang, W., Hew, K.F., Fryer, L.K., 2022. Chatbots for language learning—are they really useful? a systematic review of chatbot-supported language learning. J. Comp. Assisted Learn. 38 (1), 237–257. https://doi.org/10.1111/jcal.12610.

Jasin, J., Ng, H.T., Atmosukarto, I., Iyer, P., Osman, F., Wong, P.Y.K., Pua, C.Y., Cheow, W.S., 2023. The implementation of chatbot-mediated immediacy for synchronous communication in an online chemistry course. Educ. Inf. Technol. 28, 10665–10690. https://doi.org/10.1007/s10639-023-11602-1.

Jeon, J., 2023. Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. Comput. Assist. Lang. Learn. 36 (7), 1338–1364. https://doi.org/10.1080/09588221.2021.1987272.

Jeon, J., 2024. Exploring AI chatbot affordances in the EFL classroom: young learners' experiences and perspectives. Comput. Assist. Lang. Learn. 37 (1–2), 1–26. https://doi.org/10.1080/09588221.2021.2021241.

Jia, J., Chen, W., 2008. Motivate the learners to practice English through playing with chatbot CSIEC [Paper presentation]. In: The 3rd International Conference on

Technologies for E-Learning and Digital Entertainment, Edutainment 2008, Nanjing, China. https://doi.org/10.1007/978-3-540-69736-7_20.
Jia, J., Ruan, M., 2008. Use chatbot CSIEC to facilitate the individual learning in English instruction: a case study [Paper presentation]. In: The 9th International Conference on Intelligent Tutoring Systems, ITS 2008, Montreal, Canada. https://doi.org/10.1007/978-3-540-69132-7_84.
Jinowat, N., Unapang, J., Kangsawiwat, T., Sapai, P., 2024. Fostering critical writing excellence: leveraging google Gemini and collaborative online writing with Google Docs [paper presentation]. In: The 12th International Conference on Learning Technology for Education Challenges (LTEC 2024), Kaohsiung, Taiwan. https://doi.org/10.1007/978-3-031-61678-5_25.
Kadir, R.A., Zulkifli, M.F., Tiun, S.B., Lakulu, M.M., Jusoh, S., Faudzi, A.F.A., 2024. EduChat: AI-powered chatbot with personalized engagement for online learning. In: Arai, K. (Ed.), Intelligent Systems and Application. Springer, pp. 589–597. https://doi.org/10.1007/978-3-031-47715-7_40.
Kemelbekova, Z., Degtyareva, X., Yessenaman, S., Ismailova, D., Seidaliyeva, G., 2024. AI in teaching English as a foreign language: effectiveness and prospects in Kazakh higher education. XLinguae 17 (1), 69–83. https://doi.org/10.18355/XL.2024.17.01.05.
Kerly, A., Hall, P., Bull, S., 2007. Bringing chatbots into education: towards natural language negotiation of open learner models. Knowl. Base Syst. 20 (2), 1–9. https://doi.org/10.1016/j.knosys.2006.11.014.
Khalifa, M., Albadawy, M., 2024. Using artificial intelligence in academic writing and research: an essential productivity tool. Comput. Method. Program. Biomed. Update 5, 1–11. https://doi.org/10.1016/j.cmpbup.2024.100145.
Kim, A., Su, Y., 2024. How implementing an AI chatbot impacts Korean as a foreign language learners' willingness to communicate in Korean. System 122, 1–12. https://doi.org/10.1016/j.system.2024.103256.
Krashen, S.D., 1982. Principles and Practice in Second Language Acquisition. Pergamon Press.
Labadze, L., Grigolia, M., Machaidze, L., 2023. Role of AI chatbots in education: systematic literature review. Int. J. Educat. Tech. High. Educat. 20, 1–17. https://doi.org/10.1186/s41239-023-00426-1.
Lantolf, J.P., Thorne, S.L., 2006. Sociocultural Theory and the Genesis of Second Language Development. Oxford University Press.
Lee, K.A., Lim, S.-B., Nagarajan, S., 2023. A study of the effectiveness of English speaking of teachable agent using AI chatbot [Paper Presentation]. In: The 15th International Conference on Agents and Artificial Intelligence (ICAART), Lisbon, Portugal. https://doi.org/10.5220/0011730300003393.
Li, H., Wang, Y., Luo, S., Huang, C., 2024a. The influence of GenAI on the effectiveness of argumentative writing in higher education: evidence from a quasi-experimental study in China. J. Asian Pub. Pol. 1–26. https://doi.org/10.1080/17516234.2024.2363128.
Li, W., Chen, X., Huang, L., 2024b. Promoting second language writing through technology-driven multimodal text feedback. Innovat. Lang. Learn. Teach. 18 (5), 462–479. https://doi.org/10.1080/17501229.2024.2315104.
Lin, C.-J., Mubarok, H., 2021. Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. Educ. Technol. Soc. 24 (4), 16–35.
Lin, M.P.-C., Chang, D., 2020. Enhancing post-secondary writers' writing skills with a chatbot: a mixed-method classroom study. Educ. Technol. Soc. 23 (1), 78–92.
Lin, M.P.-C., Chang, D., 2023. CHAT-ACTS: a pedagogical framework for personalized chatbot to enhance active learning and self-regulated learning. Comput. Educ. Artif. Intell. 5, 1–8. https://doi.org/10.1016/j.caeai.2023.100167.
Liu, C.-C., Chen, W.-J., Lo, F.-y., Chang, C.-H., Lin, H.-M., 2024. Teachable Q&A agent: the effect of chatbot training by students on reading interest and engagement. J. Educ. Comput. Res. 62 (4), 1122–1154. https://doi.org/10.1177/07356331241236467.
Liu, C.-C., Liao, M.-G., Chang, C.-H., Lin, H.-M., 2022. An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. Comput. Educ. 189, 1–16. https://doi.org/10.1016/j.compedu.2022.104576.
Liu, M., Reinders, H., 2025. Do AI chatbots impact motivation? Insights from a preliminary longitudinal study. System 128, 1–12. https://doi.org/10.1016/j.system.2024.103544.
Mariani, M., Hashemi, N., Wirtz, J., 2023. Artificial intelligence empowered conversational agents: a systematic literature review and research agenda. J. Bus. Res. 161, 1–23. https://doi.org/10.1016/j.jbusres.2023.113838.
Maya, S., Camacho, M., 2024. Leveraging AI-powered mobile learning: a pedagogically informed framework. Comput. Educ. Artif. Intell. 7, 1–12. https://doi.org/10.1016/j.caeai.2024.100276.
McGrath, C., Farazouli, A., Cerratto-Pargman, T., 2024. Generative AI chatbots in higher education: a review of an emerging research area. High. Educ. 1–17. https://doi.org/10.1007/s10734-024-01288-w.
Naik, D., Naik, I., Naik, N., 2024. Applications of AI chatbots based on generative AI, large language models and large multimodal models. In: Naik, N., Jenkins, P., Prajapat, S., Grace, P. (Eds.), Contributions Presented at the International Conference on Computing, Communication, Cybersecurity and AI, July 3–4, 2024. Springer, London, UK, pp. 668–690. https://doi.org/10.1007/978-3-031-74443-3_39.
Okonkwo, C.W., Ade-Ibijola, A., 2021. Chatbots applications in education: a systematic review. Comput. Educ. Artif. Intell. 2, 1–10. https://doi.org/10.1016/j.caeai.2021.100033.
Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hrobjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo- Wilson, E., McDonald, S., et al., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst. Rev. 10, 1–11. https://doi.org/10.1186/s13643-021-01626-4.
Pan, M., Guo, K., Lai, C., 2024. Using artificial intelligence chatbots to support English-as-a-foreign language students' self-regulated reading. RELC J. 1–13. https://doi.org/10.1177/00336882241264030.
Pan, M., Lai, C., Guo, K., 2025. Effects of GenAI-empowered interactive support on university EFL students' self-regulated strategy use and engagement in reading. Internet High Educ. 65, 1–15. https://doi.org/10.1016/j.iheduc.2024.100991.
Picard, R.W., 2003. Affective computing: challenges. Int. J. Hum. Comput. Stud. 59 (1–2), 55–64. https://doi.org/10.1016/S1071-5819(03)00052-1.
Polakova, P., Klimova, B., 2024. Implementation of AI-driven technology into education – a pilot study on the use of chatbots in foreign language learning. Cogent Educ. 11 (1), 1–15. https://doi.org/10.1080/2331186X.2024.2355385.
Polyzi, P., Moussiades, L., 2023. An artificial vocabulary learning assistant. Educ. Inf. Technol. 28, 16431–16455. https://doi.org/10.1007/s10639-023-11810-9.
Reinders, H., Benson, P., 2017. Research agenda: language learning beyond the classroom. Lang. Teach. 50 (4), 561–578. https://doi.org/10.1017/S0261444817000192.
Reinders, H., White, C., 2016. 20 years of autonomy and technology: how far have we come and where to next? Lang. Learn. Technol. 20 (2), 143–154 https://doi.org/10125/44466.
Ruan, S., Willis, A., Xu, Q., Davis, G.M., Jiang, L., Brunskill, E., Landy, J.A., 2019. BookBuddy: turning digital materials into interactive foreign language lessons through a voice chatbot [Paper presentation]. In: The Sixth (2019) ACM Conference on Learning @ Scale, Chicago, IL, USA. https://doi.org/10.1145/3330430.333364.
Shafiee Rad, H., 2025. Reinforcing L2 reading comprehension through artificial intelligence intervention: refining engagement to foster self-regulated learning. Smart Learn. Environ. 12, 1–28. https://doi.org/10.1186/s40561-025-00377-2.
Shazly, R.E., 2021. Effects of artificial intelligence on English speaking anxiety and speaking performance: a case study. Expert Syst. 38, 1–15. https://doi.org/10.1111/exsy.12667.
Silitonga, L.M., Hawanti, S., Aziez, F., Furqon, M., Zain, D.S.M., Anjarani, S., Wu, T.-T., 2023. The impact of AI chatbot-based learning on students' motivation in English writing classroom [Paper Presentation]. In: The 6th International Conference on Innovative Technologies and Learning (ICITL 2023), Porto, Portugal.
Son, J.-B., Ružić, N.K., Philpott, A., 2023. Artificial intelligence technologies and applications for language learning and teaching. J. China Comput.-Assist. Language Learn. 1–19. https://doi.org/10.1515/jccall-2023-0015.
Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., Olson, C.B., 2024. Comparing the quality of human and ChatGPT feedback of students' writing. Learn. InStruct. 91, 1–15. https://doi.org/10.1016/j.learninstruc.2024.101894.
Su, Y., Lin, Y., Lai, C., 2023. Collaborating with ChatGPT in argumentative writing classrooms. Assess. Writ. 57, 1–11. https://doi.org/10.1016/j.asw.2023.100752.
Tabuchi, K., Kobayashi, S., Fukuda, S.T., Nakagawa, Y., 2024. A case study on reducing language anxiety and enhancing speaking skills through online conversation lessons. Tech. Language Teach. Learn. 6 (3), 1–21. https://doi.org/10.29140/tltl.v6n3.1497.
Tai, T.-Y., Chen, H.H.-J., 2023. The impact of Google Assistant on adolescent EFL learners' willingness to communicate. Interact. Learn. Environ. 31 (3), 1485–1502. https://doi.org/10.1080/10494820.2020.1841801.
Tai, T.-Y., Chen, H.H.-J., 2024. Improving elementary EFL speaking skills with generative AI chatbots: exploring individual and paired interactions. Comput. Educ. 220, 1–14. https://doi.org/10.1016/j.compedu.2024.105112.
Terblanche, N.H.D., 2024. Smooth talking: generative versus scripted coaching chatbot adoption and efficacy comparison. Int. J. Hum. Comput. Interact. 1–14. https://doi.org/10.1080/10447318.2024.2423125.
Tolstykh, O.M., Oshchepkova, T., 2024. Beyond ChatGPT: roles that artificial intelligence tools can play in an English language classroom. Discov. Artific. Intellig. 4 (1), 1–15. https://doi.org/10.1007/s44163-024-00158-9.
Tran, T.T.T., 2025. Enhancing EFL writing revision practices: the impact of AI- and teacher-generated feedback and their sequences. Educ. Sci. 15 (2), 1–22. https://doi.org/10.3390/educsci15020232.
Tudino, G., Qin, Y., 2024. A corpus-driven comparative analysis of AI in academic discourse: investigating ChatGPT-generated academic texts in social sciences. Lingua 312, 1–20. https://doi.org/10.1016/j.lingua.2024.103838.
Vanjani, M., Aiken, M., Park, M., 2020. Using a multilingual chatbot for foreign language practice. Iss. Inform. Syst. 21 (2), 85–92. https://doi.org/10.48009/2_iis_2020_85-92.
Vygotsky, L.S., 1978. Mind in Society: Development of Higher Psychological Processes. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4.
Wang, C., Li, Z., Bonk, C., 2024. Understanding self-directed learning in AI-Assisted writing: a mixed methods study of postsecondary learners. Comput. Educ. Artif. Intell. 6, 1–10. https://doi.org/10.1016/j.caeai.2024.100247.
Wang, C., Wang, Y., Zou, B., 2024a. Revolutionising EFL pedagogy: innovative strategies for integrating GAI (ChatGPT) into language teaching. J. Lang. Teach. 4 (1), 1–7. https://doi.org/10.54475/jlt.2024.004.
Wang, C., Zou, B., Du, Y., Wang, Z., 2024b. 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 127, 1–16. https://doi.org/10.1016/j.system.2024.103533.
Wang, Y., Xue, L., 2024. Using AI-driven chatbots to foster Chinese EFL students' academic engagement: an intervention study. Comput. Hum. Behav. 159, 1–8. https://doi.org/10.1016/j.chb.2024.108353.
Wiboolyasarin, W., Wiboolyasarin, K., Kamonsawad, R., Jinowat, N., 2024a. Enhancing collaborative writing with Grammarly-mediated feedback in asynchronous wiki

environments: an investigation of group size. Engl. Teach. Learn. https://doi.org/10.1007/s42321-024-00186-9.
Wiboolyasarin, W., Wiboolyasarin, K., Suwanwihok, K., Jinowat, N., Muenjanchoey, R., 2024b. Synergizing collaborative writing and AI feedback: an investigation into enhancing L2 writing proficiency in wiki-based environments. Comput. Educ. Artif. Intell. 6, 1–10. https://doi.org/10.1016/j.caeai.2024.100228.
Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Boonyakitanont, P., Jinowat, N., 2024c. Designing chatbots in language classrooms: an empirical investigation from user learning experience. Smart Learn. Environ. 11 (1), 1–25. https://doi.org/10.1186/s40561-024-00319-4.
Winkler, R., Söllner, M., 2018. Unleashing the potential of chatbots in education: a state-of-the-art analysis [Paper presentation]. In: The 78th Annual Meeting of the Academy of Management, Chicago, IL, USA. https://doi.org/10.5465/AMBPP.2018.15903abstract.
Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., Drachsler, H., 2021. Are we there yet? - a systematic literature review on chatbots in education. Front. Artif. Intellig. 4, 1–18. https://doi.org/10.3389/frai.2021.654924.
Wu, X.-Y., 2024. Artificial Intelligence in L2 learning: a meta-analysis of contextual, instructional, and social-emotional moderators. System 126, 1–20. https://doi.org/10.1016/j.system.2024.103498.
Yahaya, S.N., Bakar, M.H., Jabar, J., Abdullah, M.M., Segaran, Y., 2024. Evaluating students acceptance of AI chatbot to enhance virtual collaborative learning in Malaysia. Int. J. Sustain. Dev. Plann. 19 (1), 209–219. https://doi.org/10.18280/ijsdp.190119.
Yang, H., Kim, H., Lee, J., Shin, D., 2022. Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. ReCALL 34 (3), 327–343. https://doi.org/10.1017/S0958344022000039.
Ye, Y., Deng, J., Liang, Q., Liu, X., 2022. Using a smartphone-based chatbot in EFL learners' oral tasks. Int. J. Mob. Blended Learn. (IJMBL) 14 (1), 1–17. https://doi.org/10.4018/IJMBL.299405.
Yekollu, R.K., Bhimraj Ghuge, T., Sunil Biradar, S., Haldikar, S.V., Farook Mohideen Abdul Kader, O., 2024. AI-driven personalized learning paths: enhancing education through adaptive systems. In: Asokan, R., Ruiz, D.P., Piramuthu, S. (Eds.), Smart Data Intelligence. Springer, pp. 507–517. https://doi.org/10.1007/978-981-97-3191-6_38.
Yin, J., Goh, T.-T., Hu, Y., 2024. Interactions with educational chatbots: the impact of induced emotions and students' learning motivation. Int. J. Educat. Tech. High. Educat. 21, 1–23. https://doi.org/10.1186/s41239-024-00480-3.
Yin, J., Goh, T.-T., Yang, B., Xiaobin, Y., 2021. Conversation technology with micro-learning: the impact of chatbot-based learning on students' learning motivation and performance. J. Educ. Comput. Res. 59 (1), 154–177. https://doi.org/10.1177/07356331209520.
Yuan, Y., 2024. An empirical study of the efficacy of AI chatbots for English as a foreign language learning in primary education. Interact. Learn. Environ. 32 (10), 6774–6789. https://doi.org/10.1080/10494820.2023.2282112.
Yuan, L., Liu, X., 2025. The effect of artificial intelligence tools on EFL learners' engagement, enjoyment, and motivation. Comput. Hum. Behav. 162, 1–10. https://doi.org/10.1016/j.chb.2024.108474.
Zhang, C., Meng, Y., Ma, X., 2024a. Artificial intelligence in EFL speaking: impact on enjoyment, anxiety, and willingness to communicate. System 121, 1–14. https://doi.org/10.1016/j.system.2024.103259.
Zhang, L., Eichmann-Kalwara, N., 2019. Mapping the scholarly literature found in Scopus on "research data management": a bibliometric and data visualization approach. J. Librarianship Scholar. Commun. 7 (1), 1–19. https://doi.org/10.7710/2162-3309.2266.
Zhang, R., Zou, D., Cheng, G., 2024b. Chatbot-based learning of logical fallacies in EFL writing: perceived effectiveness in improving target knowledge and learner motivation. Interact. Learn. Environ. 32 (9), 5552–5569. https://doi.org/10.1080/10494820.2023.2220374.
Zhang, Z., Huang, X., 2024. The impact of chatbots based on large language models on second language vocabulary acquisition. Heliyon 10 (3), e25370. https://doi.org/10.1016/j.heliyon.2024.e25370, 1–13.
Zheng, S., 2024. The effects of chatbot use on foreign language reading anxiety and reading performance among Chinese secondary school students. Comput. Educ. Artif. Intell. 7, 1–14. https://doi.org/10.1016/j.caeai.2024.100271.
Zimmerman, B.J., 2002. Becoming a self-regulated learner: an overview. Theory Into Pract. 41 (2), 64–70. https://doi.org/10.1207/s15430421tip4102_2.